The Dissertation Committee for Mohammad Majidi-Qadikolai
certifies that this is the approved version of the following dissertation:

Transmission Expansion Planning: Computational challenges toward real-size networks

Committee:

Ross Baldick, Supervisor

Surya Santoso

Ari Arapostathis

James Eric Bickel

Constantine Caramanis

Transmission Expansion Planning: Computational challenges toward real-size networks

by

Mohammad Majidi-Qadikolai,

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2017

To my mother and father,

whose unconditional love, support and sacrifice have lit my life.

Acknowledgments

Writing this section doesn't look hard, but it really is . . .

It is hard (if possible) to list all names. I would like to thank my supervisor, committee members, colleagues in energy systems, ECE staff, friends, and family for all their help and support. I couldn't reach to this point without your help. I am very lucky for having all of you in my life . . .

Thank you all!

Transmission Expansion Planning: Computational challenges toward real-size networks

Publication No. ______________

Mohammad Majidi-Qadikolai, Ph.D.
The University of Texas at Austin, 2017

Supervisor: Ross Baldick

The importance of the transmission network for supplying electricity demand is undeniable, and Transmission Expansion Planning (TEP) studies is key for a reliable power system. Due to increasing sources of uncertainty such as more intermittent energy resources, mobile and controllable demands, and fast technology improvements for PVs and energy storage devices, the need for using systematic ways for solving this complex problem is increased. One of the main barriers for deploying optimization-based TEP studies is computationally intractability, which is the main motivation for this research.

The aim of this work is to investigate the computational challenges associated with systematic TEP studies for large-scale problems, and develop algorithms to improve computational performance. In the first step, we investigate the impact of adding security constraints (as NERC standard requirement) into TEP optimization problem, and develop the Variable Contingency

List (VCL) algorithm to pre-screen security constraints to only add those that may affect the feasible region. It significantly decreases the size of the problem compared to considering all security constraints. Then, we evaluate the impact of the size of candidate lines list (number of binary variables) on TEP, and developed a heuristic algorithm to decrease the size of this list.

In the next step, we integrate uncertainties into the TEP optimization problem and formulate the problem as a two-stage stochastic program. Adding uncertainties increases the size of the problem significantly. It leads us to develop a three-level filter that introduces important scenario identification index ($ISII$) and similar scenario elimination (SSE) technique to decrease the number of security constraints in stochastic TEP in a systematic and tractable way.

We then investigate the scalability of the stochastic TEP formulation. We develop a configurable decomposition framework that allows us to decompose the original problem into subproblems that can be solved independently and in parallel. This framework can benefit from using both progressive hedging (PH) and Benders decomposition (BD) algorithms to decompose and parallelize a large-scale problem both vertically and horizontally. We have also developed a bundling algorithm that improves the performance of PH algorithm and the overall performance of the framework.

We have implemented our work on a reduced ERCOT network with more than 3000 buses to demonstrate the practicality of the proposed method in this work for large-scale problems.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

The transmission network is the backbone of the electric power system. Increasing penetration of renewable resources, energy storage devices, mobile and flexible demand, along with new public policies such as the "Clean Power Plan" makes the future much more uncertain for Transmission Expansion Planning (TEP). As the transmission network is a monopoly infrastructure, and in jurisdictions such as Electric Reliability Council of Texas (ERCOT) its investment and operation costs are distributed between all electricity users in the region, it is critical to expand and operate this network at minimum cost while keeping a high level of reliability. Transmission Expansion Planning is the process of deciding which equipment should be selected, where it should be installed, and when is the best time to install it. In dynamic TEP, planning is done for multi-stages, in which a decision about the best time to install is also made [103].

Villasana et al in [113] provides a hierarchy of three questions that should be answered in transmission planning:

a) What new facilities should be installed so that future operation will not be limited by transmission capacity?

b) What new transmission facilities can be economically justified versus the higher operation costs if new facilities were not installed?

c) What new Generation sites can be justified versus new transmission facilities or higher operation costs?

These three questions specify main components of the objective function in TEP. In question a), the objective function is to invest in the transmission network as much as we need to supply all demand and the impact of power system operation cost on TEP is ignored. It is sometimes called reliability planning, in which the main concern is satisfying network reliability criteria. Unit operation set points are mainly defined based on experience or least cost. In the case of using lower operating cost units as much as possible, we will have the least operation cost but we may need to invest highly in transmission expansion, posing the question of whether the investment is cost-effective. In the next hierarchy level (question b), the impact of operation cost on decision making for TEP is considered, which means it might be economical to dispatch some expensive power plants to supply demand instead of building some new transmission lines to dispatch all cheap power plants. The second question provides a better modeling property compared to the first one as it economically adjusts investment and operation costs, but it is computationally more expensive. In question c), which has the highest rank in the hierarchy, not only the impact of operation cost but also the impact of investment in generation sector on TEP is evaluated. In other words, it might be economical to invest on the generation side (for example building new power plants close to demand) instead of the transmission side to supply the demand. It provides a better expansion plan (from economical perspective); however it is much more computationally expensive, and planners would need to have the authority to make decisions about the location/capacity of new power plants. Since generation expansion decisions are usually made by individual private investors in vertically unbundled electricity industries, the consideration of generation investment may be beyond the control of transmission planners. In this dissertation, we try to answer the second question, and we assume we know the location and capacity of future generation units (with uncertainties). In principal, generation expansion could be added to the formulation.

1.2 Factors Affecting Transmission Expansion Planning

TEP studies are performed in different time-scales i.e near-term (for less than or equal to five years) and long-term (for more than ten years), and for each time-scale different parameters with different level of details are considered. Main parameters that affect TEP are categorized into four main groups namely environmental issues, legal issues, uncertainties, and network modeling, and these are explained briefly in the following:

Environmental issues: Environmental concerns/limitations may directly affect transmission planning especially for line routing in particular areas such as:

- Regions with wildlife and endangered species,
- Wetlands,
- National parks, historic areas, military areas etc.

Furthermore, there are some environmental concerns that indirectly affect transmission planning such as:

- Limits on pollution generated by power plants in different areas,
- Access to water resources necessary for building and operating power plants and etc.

These will directly affect the generation expansion and indirectly affect transmission planning as there is a dependency between generation and transmission expansion planning.

Legal issues: Policy makers can affect TEP in several different ways such as:

- Who should pay for new transmission lines? For example, all entities that are connected to the network or just those who benefit from the line?
- What should be the transmission usage tariffs?
- What are electricity market price caps?

Uncertainties: There are several uncertainties that affect TEP, and we should try to address them during the planning stage. They mainly can be categorized as micro and macro uncertainties:

- Macro uncertainties such as future changes in economic growth, market rules, carbon emission issues, fuel price, generation mix/location and capacity, technology revolutions etc.

- Micro uncertainties such as load and intermittent resource variations, availability of power plants and transmission lines in real time, market price, behavior of market participants etc.

The macro uncertainty may be well represented by probability distributions, and an expected cost framework may be sufficient to capture main issues. The macro uncertainties may not have well-defined probability distributions, and risk may be much more important in this context, motivating approaches such as robust optimization [13, 105]. In this work, we will primarily consider uncertainties that have well-defined probability distributions. When the distribution, either the family or the member, is not well-defined or known, other methods are required to address the additional uncertainties (see [46, 38] for more detail).

Power system modeling: There are different models for different network components and operation such as:

- Steady-state power flow formulation: It can be divided into three main categories i.e. transportation model in which only the first Kirchhoff's law is satisfied, the DC model that satisfies both both first and second Kirchhoff's laws, ignoring network losses and reactive power requirements, and the AC model which is the most accurate model for power system steady-state modeling and considers network losses and reactive power requirements as well as the first and the second Kirchhoff's laws. There are some hybrid models that are mainly derived form one of these three main models such as DC model with linear approximation of network losses or linearized AC model with loss and reactive power modeling.

- Transmission network model: Transmission network can be modeled as non-controllable or controllable. In the non-controllable model, the topology of the network is fixed, and in controllable model, it is possible to use switching, phase shifters, FACTS devices, special protection schemes, etc to control and manage flow directions in the gird.
- Generation model: There are several parameters that affect a power plant operation i.e. its maximum and minimum capacity limits, ramp rate capability, and some limits that are driven by specific generation technologies like total energy limit for hydro power plants (based on their reservoir capacity), etc.
- Demand model: There are two different ways to model load i.e. elastic or inelastic. In the elastic model, demand can be controlled with different signals such as the market price, but in the inelastic model, demand is modeled as a fixed quantity that should be supplied, if possible, and only curtailed in case of scarcity.
- Operation states: Normal and under contingency are two different

types of operation states that can be evaluated in power system analysis (for both steady-state and transient analysis).

- Market model: There are several different aspects in market modeling like ideal versus real markets, day-ahead vs. real-time, etc. that may affect system operation costs and TEP.

Selecting different models affects the accuracy of results and computational time required to solve the problem.

1.3 Systematic Transmission Expansion Planning

Based on above mentioned significant parameters, TEP is a multi-dimensional and very complex problem. The question is how to model/formulate all of these parameters, and a harder question will be how to solve that problem for large-scale networks. Making assumptions and simplifications seems inevitable, and we seek to do so in a way that does not fundamentally invalidate the analysis. Environmental and legal issues mostly can be considered in near-term TEP/line design stage, and can be partially addressed in developing candidate lines for long-term TEP. Therefore, we can model their impacts outside of TEP optimization formulation and thereby significantly reduce TEP problem size. Uncertainties can be captured by developing different possible scenarios and using either heuristic methods or stochastic programing to solve or by developing uncertainty boundaries and using robust optimization for problem formulation. How to model the power system and integrate uncertainties is categorized by [113] into five different stages.

Stage I: considering all quantities deterministic (future load, generation, fuel price and etc), static model (one planning horizon), single operation condition (normal operation), all variables as continuous (continuous line capacity for expansion);

Stage II: deterministic quantities, static model, single operation condition, mixed-integer problem (MIP) statement (binary decision variables for building transmission lines);

Stage III: deterministic quantities, static model, multi operation conditions (normal and under contingency operation states), MIP statement;

Stage IV: deterministic quantities, dynamic model (multi-planning horizons), multi operation conditions, MIP statement;

Stage V: stochastic quantities (uncertainties in load, generation, fuel price, and etc), dynamic model, multi operation conditions, MIP statement.

By moving from stage I to stage V, the model will be more accurate and close to reality, but much more complicated and challenging to solve. By using DC model, stage I represents a continues optimization problem. Adding integer variables makes it a Mixed-Integer Programming (MIP) problem in stage II. Stage III adds contingency analysis into TEP that significantly increases problem size and can easily make TEP optimization problem intractable. TEP moves from static to dynamic in stage IV that increases the number of binary variables in the optimization formulation, and TEP is modeled as stochastic dynamic TEP in stage V. In this dissertation, we move from stage II toward stage V (but not at the same order proposed by [113]) and try to develop algorithms that make it possible to solve large-scale problems. For distribution grid planning and investment decision making please see [79] and [78].

1.4 Layout of this dissertation

This dissertation is organized as follows:

In chapter 1, an introduction about transmission expansion planning is provided and main factors that affect TEP are discussed. It is followed by discussion of stages in systematic TEP for adding details in the modeling.

In chapter 2, literature on TEP along with NERC's reliability standard for transmission planning are briefly reviewed. Deterministic TEP for multiple operation states with $N-1$ contingency analysis is modeled. A constraint screening algorithm is developed to screen reliability constraints and select a subset of lines for contingency analysis such that their outage will cause overload in other lines in the network.

In chapter 3, the impact of uncertainties on TEP is explicitly modeled by developing a two-stage stochastic optimization formulation. The main focus of this chapter is to develop an algorithm to reduce the number of candidate lines when the initial candidate line list is very large. The developed heuristic method will decrease the number of binary decision variables and improve the performance of MIP for TEP. Contingency analysis is not integrated in TEP in this chapter.

In chapter 4, contingency analysis is added to stochastic TEP optimization problem (compared to chapter 3) that significantly increases the size of the problem. Therefore, a framework is designed that solves the problem iteratively to improve computational performance. A three level filter is designed to select a subset of reliability constraints in each iteration and gradually increase the size of the problem. The VCL algorithm developed in chapter 2 is part of this filter. The result of this algorithm will be an upper bound for TEP problem, and a lower bound can be found using branch and bound technique to quantify the quality of results.

In chapter 5, a general decomposition framework is developed to solve large-scale TEP problems as the developed method in chapter 4 became intractable for large-scale problems. It can benefit from both Benders decomposition and Progressive Hedging algorithms for the same problem. A bundling algorithm is developed to improve the convergence of the progressive hedging algorithm. The proposed method in chapters 2 and 4 can be used to solve each subproblem.

In chapter 6, different factors affecting the performance of the proposed framework in Chapter 5 are investigated, and two case studies are used to demonstrate the capabilities of this framework.

In chapter 7, a summary of main findings of this research is provided. This chapter concludes with a discussion of different paths for future work in this area.

Chapter 2

Reliability constraint screening for the TEP optimization problem [1]

[1]Mohammad Majidi-Qadikolai and Ross Baldick. Integration of N-1 contingency analysis with systematic transmission capacity expansion planning: ERCOT case study. IEEE Transactions on Power Systems, 31(3):2234-2245, May 2016. Authors had equal contributions.

Nomenclature

Sets and Indices:

N_b: Set of buses with index *i, k, n, r (for reference bus)*

N_g: Set of all generators with index g

N_l: Set of all lines (existing and candidate) with index *l, m*

N_o: Set of all existing lines with index *l, m*

N_n: Set of all candidate lines with index *l, m*

N_u: Set of all existing lines and selected candidate lines with index *l, m*

CLL_o: Original candidate lines list

CLL_u: Set of updated candidate lines with index *l, m*

L_k: Set of lines connected to bus k

G_k: Set of all generators connected to bus k with index k

W_k: Set of wind generators connected to bus k with index k

Φ_l: Set of lines with violated post-contingency flows under outage of line l

N_s^t: Set of system operation states in load block t with index c ($c = 1$ represents normal operating condition)

T: Set of load blocks with index t

Ω: Set of all scenarios with index ω

$|\ |$: Size of a set

t: Superscript for different load blocks

ω: Superscript for different scenarios

Parameters:

q_i: Per MWh load curtailment penalty at bus i

γ_i: Per MWh wind curtailment penalty at bus i

Co_g: Per MWh operation cost for generator g

ζ_l: Annual cost of line l construction

η: Line loading threshold for monitoring purpose

d_i: Demand at bus i

P^{ω}: Probability of scenario ω

B: Diagonal matrix of line admittance

Y: Reduced admittance matrix (column and row related to reference bus are removed)

Ψ: Reduced bus-branch incidence matrix (row related to reference bus is removed)

P_g^{max}: Maximum capacity of generator g

P_g^{min}: Minimum capacity of generator g

f_l^{max}: Maximum capacity of line l

f_l^{min}: Minimum capacity of line l

M_l: Big M is a large positive number for line l

C: Matrix of contingencies that specifies the status of lines under different contingencies (1 for in service and 0 for out of service lines) with index c

$PTDF$: Power transfer distribution factor

$LODF$: Line outage distribution factor

$LCDF$: Line Closure Distribution Factor

$\Gamma_{m,l}$: Magnitude of violation in flow of line m when line l is on outage

CII_l: Contingency identification index for outage of line l

α: Line capacity modification factor for short-term capacity limits

χ: Reduced Z-bus matrix (inverse of Y)

$Flag$: Is set to 1 if any load or generator is connected to an island bus in the base case, and is set to 0 otherwise.

Variables:

x_l: Binary decision variable for line l

$r_{i,c}$: MW load curtailment at bus i under operation state c

CW_i: Aggregated MW wind curtailment at bus i

p_g: Output power of generator g

$f_{l,c}$: Power flow in line l under operation state c

$\theta_{i,c}$: Voltage angle at bus i under operation state c . $\Delta\theta_{l,c}$ is voltage angle difference across line l under operation state c, $\Delta\theta_{l,c}$= $\theta_{k,c}$-$\theta_{n,c}$ for line l from bus k to bus n.

2.1 Introduction

From the optimization formulation perspective, TEP is a large scale, non-convex and nonlinear optimization problem. Using linear approximation of AC power flow equations is one of the most popular simplifications for modeling non-linear power flow equations in high level TEP. The accuracy of linear approximation of power flow equations (DC model) is evaluated in [37, 111, 9, 93]. In [37], authors defined "overload network" to model overloads in different corridors in existing network and make decision about new line requirements. The designed network with linear approximation is tested with AC power flow equations, and if there is no occurrence of an overload in AC analysis then this approximation passed the AC test successfully. In [111], the authors compared the results of AC and DC power flow results for the IEEE 300-bus system, and showed the error between DC and AC results will be less than 5% when the assumptions of DC power flow are satisfied. Authors in [9] compared the sensitivity analysis in power systems with DC and AC models, and demonstrated that it provides a relatively reliable approximation of the behavior of the system. In [93], the authors showed that locational marginal prices (LMPs) that drive the economic analysis of power system operation will not be significantly affected when AC model is approximated with DC model.

In [113] and [37], transmission planing is formulated as a simple linear programming (LP) problem with continuous decision variables. In [113], they proposed a LP method with continuous variables for optimal transmission planning by minimizing load curtailment. As transmission line capacity is lumpy, considering capacity to be a continuous variable is not accurate. In [112], the author proposed a mixed integer programming (MIP) formulation using binary decision variables for selecting new lines with DC power flow approximation. This method is more accurate in representing new line capacities, but their formulation is not computationally efficient.

Kirchoff's second law can be represented with two inequalities in a mixed integer disjunctive model, each related to one possible flow direction [8]. This technique increases the number of constraints and provides better conditioning properties by tightening constraints. The authors of [8] also used GRASP meta-heuristic method to provide an upper bound feasible solution. In [2], power network losses are integrated into TEP optimization problem by using piecewise linear loss functions for each line. It provides more accurate power system model for planning purpose while preserving linearity, and may affect the selected expansion plan for networks with relatively high losses such as systems with long transmission lines. However, the simulation time for this case is increased around five times compared to the case without losses. This huge extra computational burden should be added to the model if it is expected to have a significant impact on selected plans (based on average network losses in the area of study). A detailed analysis on the impact of line loss modeling on AC power flow approximation is given by [21].

Benders decomposition (BD) is used in several contexts as a powerful tool for decreasing simulation time for solving large scale optimization problems. Mathematical formulation for implementing Benders decomposition for transmission and generation expansion planning is developed by EPRI in 1988 [43]. Gomory cuts are added to Benders cuts in [14] to improve the performance of BD for large scale MIP problems. To overcome the non-convexity of transmission planning problem, [103] and [102] proposed a three phase hierarchical decomposition method to find the global optimal answer. They used BD to solve each phase and transfered Benders cuts into the next phase to integrate different phases. They used a transportation model, in which the second law of Kirchoff is relaxed, and a hybrid model (transportation model for new lines and DC model for existing lines) with continuous variables (LP model) to get the global optimal results (for their approximate formulation) in the first and second phases of their hierarchical model. In the third phase, they used the DC model with discrete decision variables and Benders cuts from the first and the second stages to solve MIP optimization problem. In [94], authors considered load and wind as dependent and uncertain variables, and used a two stage stochastic model and sequential approximation technique to solve TEP optimization problems with BD. A dynamic transmission expansion planning is formulated in [84] and authors compared the performance of stochastic programming with deterministic and heuristic methods. In [86], authors evaluated the impact of different approximations on TEP with renewable portfolio standards. Authors in [85] and [87] proposed a new approach for multi-regional transmission and generation expansion planning with Benders decomposition technique, which is enhanced by developing new lower bounding constraints that increase convergence speed. They applied the model to large scale networks with a relatively large number of scenarios to capture uncertainties, and evaluated the impact of optimality gap on simulation time. To decrease computational efforts, all above mentioned references ignored $N-1$ contingency analysis in their proposed methods for transmission planning. So, there is no guarantee that selected optimal plans by these papers satisfy $N-1$ criterion.

2.1.1 Power System Adequacy and Reliability

The power system should be adequate and reliable. Based on North American Electric Reliability Corporation (NERC) definition "Adequacy is the ability of the electric system to supply the aggregate electric power and energy requirements of the electricity consumers at all times, taking into account scheduled and reasonably expected unscheduled outages of system components" and "Operating reliability is the ability of the electric system to withstand sudden disturbances such as electric short circuits or unanticipated loss of system components" [90]. In standard 51, NERC categorized system adequacy and security into four levels A-D [89]. Level A refers to system performance under normal conditions (no contingency), and in level B, system performance following the loss of a single bulk system element is evaluated. In Level C and D, system performance under loss of two or more bulk system components and extreme events are evaluated, respectively. Categories A-C should be evaluated for near-term planning (one to five years) and long-term planning (more than ten years), and category D should be considered for near-term planning only.

The power system should be planned and be operated in a way to be able to supply all loads in case of a single outage in system components (level B), which is called $N-1$ criterion [89], [27]. To satisfy this standard, system operators usually use security constrained optimal power flow (SCOPF) or security-constrained unit commitment (SCUC) to dispatch/commit power plants. Post-contingency re-dispatch [81], congestion management [63], transmission switching [106, 118, 48, 67, 107], or using FACTS devices [62, 65, 124, 123, 125] are techniques used to add flexibility to transmission operation and subsequently reduce operation costs. In [81], a new algorithm for security constrained optimal power flow (SCOPF) is proposed that considers post-contingency corrective rescheduling to decrease dispatch costs. In [48], authors applied a sensitivity analysis to the economic impact of transmission switching that shows that incremental switching benefits will decrease when the number of allowed switching operations increases. Authors in [67] added some constraints to transmission switching optimization problem to make it more practical by limiting the number of switching. They have also proposed a heuristic method to reduce the switching lines list to decrease computation time. To integrate transmission switching in system operation, authors in [107] used flow cancellation technique to model switching. They showed that this technique is faster than using binary variables to change the status of lines in topology control when the number of switching lines in limited. In [114], transmission switching is integrated with TEP, and they showed that switching can change transmission expansion plans by alleviating contingencies and decreasing power system operation costs in systems with high wind penetration, but they also ignored the impact of contingency analysis on planning.

Various researchers use either $N-1$ criterion or probabilistic approaches such as Loss of Load Probability (LOLP) or Loss of Load Expectations (LOLE) for power system adequacy and security evaluation. In [59], authors explained drawbacks of each method and evaluated the impact of considering different reliability criteria on TEP. They performed numerical analysis for the Garver 6-bus system [37] to compare the performance of these methods. The result shows that TEP with $N - 1$ criterion requires more investment compared to TEP with probabilistic approaches as it should supply the demand under all single contingencies. Loss of Load Cost (LOLC) as a reliability index is calculated for the selected plan for both cases, and LOLC for TEP with $N-1$ criterion is much less than LOLC for TEP with the probabilistic approach, showing the impact of extra investment on improving system reliability. By considering $N - 1$ criterion, the system quality and reliability indexes will be less sensitive to load variations and components' rate of outage compared to probabilistic approaches.

O'Neill et al proposed a comprehensive mathematical formulation for dynamic optimal power system planning and investment by integrating unit commitment, transmission switching, and $N - 1$ contingency analysis into a power system operation cost formulation in [92]. But as they mentioned in their paper, it is a very complex and computationally expensive model even for a very small case study; so it is not practical for large scale networks at this time. More practical formulations for TEP optimization with $N-1$ contingency analysis are formulated in [104, 54, 83, 122, 66, 70]. Rudkevich [104] proposed a nodal capacity market framework for generation and transmission expansion planning. He used the flow cancellation technique to represent a fixed list of contingencies in a reliability dispatch formulation, in which all resources are dispatched at zero costs and load shedding will be penalized at value of lost load (VOLL) price. In [54], authors proposed a three stage transmission and generation expansion planning optimization formulation with Benders decomposition technique, and considered contingency analysis for all existing and candidate lines and integrated transmission switching to alleviate violations in line flows. In [83], authors developed a probabilistic method for transmission investment by integrating security and corrective controls into operation cost estimation. Generation reserves and special protection schemes are considered as additional corrective actions that can help system stability during contingencies and decrease operation costs by reducing load shedding. This method is applied to the IEEE 24-RTS case study, and it takes more than 600 seconds to solve this case with 40 different operating conditions (which is parallelized on a machine with 12 cores and 192 GB of RAM), and they did not evaluate the performance of their proposed method for large scale systems. In [122], an iterative method for multi-stage transmission planning is proposed by integrating linear approximation of network losses into MIP formulation. $N-1$ contingency analysis is not integrated into the planning formulation, and it is checked separately in a security check sub-problem. To decrease the problem size, they suggested to limit the set of monitored lines during contingency analysis to the lines that are close to the expansion area.

Because transmission line loading depends on demand and generation dispatch, there is no guarantee that TEP with a fixed contingency list (used by [104]) will satisfy $N-1$ criterion in all conditions, while including all lines in the contingency list (used by [54] and [122]) is not necessary most of the time in real networks, as usually a single outage of only a limited number of lines will actually result in violation in flow limits on other lines. For example in ERCOT, there were only about 700 contingency constraints (out of tens of millions of possible constraints) which were binding at some time during 2013 [97]. Usually during midnight with very low load level, single outage of any line will not cause overload on other lines in most power systems. In other words, constraints related to those lines will remain passive in the optimization problem, and will not affect the feasible region and the optimal answer. Therefore, for this particular light load case we do not need to consider all or part of lines for contingency analysis, and results of OPF will be feasible for SCOPF as well. But this is not the case for all loading conditions, and important lines for contingency analysis that will contribute to forming the feasible region depend on the system loading condition and network configuration, so the key question is how to find them. In this chapter, we propose a method to screen reliability constraints and select effective lines for contingency analysis based on system conditions.

The rest of this chapter is organized as follows: in section 2.2, the proposed reliability constraints screening framework is explained, and in section 2.3 mathematical formulation of transmission expansion planning with $N-1$ contingency analysis with model performance are discussed. In section 2.4, the proposed method is implemented on two different case studies, and results are compared with the integrated model, in which all contingencies are integrated into TEP optimization problem.

2.2 The Proposed Reliability Constraints Screening Framework

2.2.1 Modeling Assumptions

As stated in chapter 1, there are several parameters that affect the selected optimal plan in TEP. It is almost impossible to model/integrate all those important parameters in TEP, and be able to solve the problem for large scale systems in a reasonable time with current machines. Therefore, we need to choose some parameters that have more significant impact on the long-term TEP. In this work, we assume legal issues and environmental constraints can be addressed outside of TEP optimization problem as they usually affect transmission candidate lines. Voltage and reactive power requirements usually have local impacts and the investment cost to address possible issues in this area is much less than investment in new transmission lines, so we can ignore them in the high-level TEP and address them in near-term planning (this assumption is valid for most transmission systems but not all, so if the system reliability is too sensitive to voltage and reactive power in a network it should be integrated into TEP formulation). Network losses can be an issue for systems with relatively long transmission lines and high losses but otherwise their impact is typically negligible. Transient stability is a critical issue in power system operation and design, and to be analyzed accurately it needs detailed network data that usually are not available during long-term planning. A less accurate evaluation can be done using typical data to make sure there is no significant reliability issue with the selected expansion plan; therefore it will not add extra computational burden on TEP optimization problem. Other assumptions in this chapter are as follows:

- Linearized power flow equations with the first and second Kirchhoff's lows (DC power flow),
- Considering a limited number of load blocks to represent load and wind variations,
- Unit commitment (UC) is approximated with optimal power flow, assuming all generation is in-service,
- The market is competitive, so power plants offer at their marginal costs and they do not exercise market power,
- Load and wind are modeled as deterministic in each load block with multiple operation load blocks to capture load and wind variations.

To satisfy "ERCOT Planning Guide, Section 4: Transmission Planning Criteria" [27], NERC's standard on transmission planning [91] and standard 51 [89], we integrate $N-1$ contingency analysis into TEP (instead of probabilistic approaches).

In most restructured electricity industries, generation capacity expansion planning (GCEP) is decentralized and private parties make their own decisions regarding the location, capacity, and type of new power plants. In this chapter, generation expansion planning is assumed deterministic and sufficient to supply demand; therefore, TEP optimization problem is solved for a given future load and generation growth. However, there are macro uncertainties in future generation expansion, and a more realistic way to model this uncertainty is to run TEP under different GCEP scenarios rather than considering deterministic future generation expansion (as will be investigated further in next chapters). Moreover, by moving from static TEP toward dynamic (multi-stage) TEP, it is possible to model uncertainties about GCEP in the future (wait-and-see) for modifying TEP. Modeling dynamic TEP significantly increases computational time for large scale systems, so it should be considered carefully.

For power system operation cost modeling, we consider preventive DC-SCOPF, in which single outage of lines and transformers are considered as contingencies for satisfying $N-1$ criterion and power plants will have the same dispatch during normal and under contingency operating states (see [6] for more discussion about preventive DC-SCOPF in transmission planning).

In long-term (10+ years) planning, corrective dispatch is usually considered for single outage of generators during the planning process ($G-1$) [54], [122]. However, in near-term transmission planning (less than 5 years), it is common to include different corrective actions for operation cost estimation to mitigate the impact of an outage and develop corrective expansion plans [91]. Compared to corrective DC-SCOPF, preventive DC-SCOPF usually results in higher operating costs as it has tighter constraints, and provides higher security margin for $N-1-1$ security condition, in which the system should withstand the second outage when the system is returned to the normal operation condition after the first outage. The higher operating cost as a result of preventive dispatch may affect the selected optimal transmission expansion plan (usually increasing investment costs). The proposed algorithm in this section can be applied to TEP with both corrective and preventive DC-SCOPF formulations.

Considering single outage of all lines as operation states will increase the size of the problem significantly and usually makes it unsolvable; therefore different techniques are used for selecting some lines for contingency analysis. Authors in [61] proposed a constraint screening method for security analysis in which SCUC formulation is replaced with an equivalent reduced-order SCUC problem to decrease computational time. In [32], contingencies are categorized as more probable and less probable. This criterion will result in selecting lines that have more possibility of outage. This will not guarantee satisfying $N-1$ criterion, and is more suitable for probabilistic approaches. Other techniques for decreasing computational time for contingency analysis include removing parallel lines from contingency list, decreasing the number of monitored lines, and selecting important lines based on expert knowledge [117]. By omitting the lines such that their outages do not cause overload on other lines, we can decrease the number of lines for contingency analysis, and increase simulation speed while still satisfying $N-1$ criterion. We propose a systematic method to automatically create contingency lists based on network configuration and loading conditions, and integrate them into the TEP problem to find the optimal transmission capacity expansion plan that satisfies $N-1$ criterion.

2.2.2 PTDF and LODF concepts and formulation

Power Transfer Distribution Factor ($PTDF$) and Line Outage Distribution Factor ($LODF$) are two factors mainly used for sensitivity analysis of flows on transmission lines. $PTDF$ is defined as a measure for sensitivity of line MW flow to a MW transfer. A MW transfer in this definition refers to injecting $1MW$ at one bus and withdrawing $1MW$ at another bus. There are different formulations for calculating $PTDF$. Some references define $PTDF$ as a MW transfer when the withdrawal bus is always the reference bus (r), so the only important parameter for calculating $PTDF$ based on this definition will be the injection point. In [109], $PTDF$ formulation is given based on this definition as follows:

$$PTDF_{l,(ir)} = (\chi_{ni} - \chi_{ki}) \times B_{l,l} \tag{2.1}$$

$$\tilde{f}_l = f_l + PTDF_{l,(ir)} \times \Delta P_i \tag{2.2}$$

where $PTDF_{l,(ir)}$ represent the sensitivity of flow on line l joining between bus n and k to the injection of power at bus i and withdrawal at the reference bus r. χ_{ni} represents the element (n, i) of the reduced Z-bus matrix χ, and $B_{l,l}$ is the susceptance of lines l. $\tilde{f}_l$ is flow on line l after injecting ΔP_i at bus k. f_l shows flow on line l before injecting power at bus i.

The Line Outage Distribution Factor ($LODF$) shows the sensitivity of flow on lines in a network when there is a change in flow of a line in that network. In other words, $LODF_{m,l}$ shows the percent of pre-outage flow on line l that will show up on line m after the outage of line l.

$$LODF_{m,l} = \frac{\Delta f_{m,l}}{f_l} \tag{2.3}$$

where $\Delta f_{m,l}$ shows change on flow on line m after the outage of line l, and f_l represents pre-outage flow on line l.

By using $PTDF$ defined in (2.1), $LODF$ can be calculated as follows [109]: Assume we inject ΔP_n at bus n and withdraw at reference bus. Suppose we specify ΔP_n as follows:

$$\Delta P_n = \frac{-f_l}{PTDF^{IN}_{l,(nr)}} \tag{2.4}$$

Superscript IN represents the case that line l is connected to the network. By injecting this power, flow on lines will change. Flow on line l and m can be calculated by equation (2.2). For line l, as shown in (2.5), $\tilde{f}_l = 0$.

$$\begin{aligned} \tilde{f}_l &= f_l + PTDF^{IN}_{l,(nr)} \times \frac{-f_l}{PTDF^{IN}_{l,(nr)}} \\ &= 0 \end{aligned} \tag{2.5}$$

Therefore, if we open line l after making the change ΔP_n to injection at bus n then flows in the rest of network will not change. After opening this line, $PTDF$ coefficients will change because of changing network topology, and we use superscript OUT to distinguish it from the case that line l was connected.

We successfully opened line l without changing flow on other lines in the network at the cost of injecting ΔP_n. Now to remove this extra injection, we inject $-\Delta P_n$ at bus n (note that line l is opened now) so the net injection at this bus will be zero. This new injection will change flow on lines in the network again. The total change in flow on line m as a result of opening line l can be calculated from the following equation:

$$\Delta f_{m,l} = PTDF^{IN}_{m,(nr)} \times \Delta P_n + PTDF^{OUT}_{m,(nr)} \times -\Delta P_n \tag{2.6}$$

By substituting (2.4) into (2.6), and reordering based on (2.3), we will have the following equation:

$$LODF_{m,l} = \frac{PTDF^{OUT}_{m,(nr)} - PTDF^{IN}_{m,(nr)}}{PTDF^{IN}_{l,(nr)}} \tag{2.7}$$

Using equation (2.7), we need to calculate $PTDF$ both before (with superscript IN) and after (with superscript OUT) opening a line which is not computationally efficient.

$PTDF$ can be defined as sensitivity of injecting and withdrawing $1MW$ power between two specific buses. This definition can be related to the previous definition of $PTDF$ (2.1) as follows:

$$PTDF_{l,(ni)} = PTDF_{l,(nr)} - PTDF_{l,(ir)} \tag{2.8}$$

where $PTDF_{l,(ni)}$ shows sensitivity of flow of line l to $1MW$ injection at bus n and withdrawal at bus i. In this equation, n and i can be any buses in the network.

NOTE: Indices m and l are exclusively used for transmission lines and indices i, n, r and k are dedicated to buses. To distinguish between bus versus line, we put the index in parenthesis if it refers to a bus. If bus n and i are two ends of line m, then we use notation $PTDF_{l,m}$ instead of $PTDF_{l,(ni)}$ for the left side of equation (2.8) for simplicity.

It is also possible to model $LODF$ as a transfer [98]. The concept is shown in Figure 2.1. A transfer $\Delta P_n = \Delta P_k$ is created on two ends of line l i.e. buses n and k. The magnitude of transfer is selected in way that $\Delta P_n = \tilde{f}_l$. In this case, flow on switches n and k will be zero (based on the first law of Kirchhoff), so opening those switches will not affect network dispatch. This is known as the flow cancellation technique, which is used by [104]. This transfer will affect flow on other lines in the network as well that can be calculated from the following equation, where we no longer use the superscripts "IN" and "OUT", and all $PTDF$s correspond to the "IN" case before opening the line.

$$\Delta f_{m,l} = PTDF_{m,l} \times \Delta P_n = PTDF_{m,l} \times \tilde{f}_l \tag{2.9}$$

Based on (2.2), post transfer flow on line l is:

$$\tilde{f}_l = f_l + \Delta f_{l,l} = f_l + PTDF_{l,l} \times \tilde{f}_l \tag{2.10}$$

$$\Delta P_n = \tilde{f}_l = \frac{f_l}{1 - PTDF_{l,l}} \tag{2.11}$$

By substituting (4.21) and (2.11) in (2.3), we will have:

$$LODF_{m,l} = \frac{PTDF_{m,l}}{1 - PTDF_{l,l}} \tag{2.12}$$

Equation (2.12) models LODF as a transfer. Direct method is another method that can be used to calculate LODF. It is similar to the transfer method, and is explained in subsection 2.3.2 with VCL algorithm.

2.2.3 The Proposed Framework

The number of new lines that should be selected and the number of operation states (normal and/or under contingency) that should be integrated into MIP have significant impacts on computational effort and simulation time. The proposed method uses these facts to solve large scale TEP with $N-1$

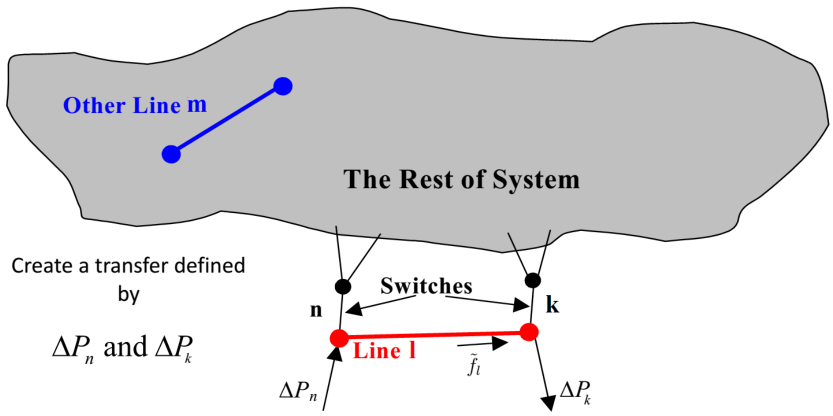


Figure 2.1: Modeling LODF as a transfer [98]

contingency analysis faster. The proposed framework can be summarized in the following steps:

Step 1: Load Input Data.

Input data includes load, generation, current and candidate transmission components etc. The base case system that contains existing lines, load, buses and generators is referred to as S_o.

Step 2: Check system islanding.

In this step, all candidate lines are tentatively modeled as being built. If there are any island buses in this system, it means those buses will not have any impact on making a decision about candidate lines so they can be safely removed from data. The base case system S_o is modified by removing these islands.

Step 3: Solve a relaxed version of TEP problem.

In this step, we solve a relaxed version of the original integrated TEP, in which all constraints related to contingency analysis are ignored (for the base case S_o). This optimization problem is much easier to solve and provides a lower bound for the original problem. The selected candidate lines in this step together with the existing network (N_o) form N_u. The updated system that contains the base case system (S_o) and newly added lines is referred to as S_u.

Step 4: Temporarily remove island buses from S_u.

As it is still possible to have island buses in system S_u and we need a system without any island buses for the next step, these buses will be removed from bus data in S_u temporarily creating a reduced system that is referred to as S_r. Island buses will not have any impact on creating contingency lists because no line is connected to them. We use S_r in step 5 to create VCL.

Step 5: Create variable contingency list (VCL).

Modified Line Outage Distribution Factor (LODF) matrix is calculated for single outage of all lines in S_r. Based on these factors, we calculate post-contingency flow in transmission lines (for DC power flow) and Contingency Identification Index (CII) for each line on outage. VCL will be created based on CII and other given parameters. If VCL is empty, it means the current network configuration satisfies $N-1$ criterion, go to step 7. Otherwise, Contingency matrix (C) will be created based on

VCL. It should be mentioned that if there are T load blocks to simulate yearly operation period, CII^t, VCL^t, and C^t should be calculated for all $t \in T$. Mathematical formulation is given in subsection 2.3.2.

Step 6: Solve TEP optimization problem.

In this step, two different options are considered for solving TEP optimization problem as follows:

- Option (i): System S_u, to which lines from step 3 are added, is used as the base case system for solving TEP with contingencies for this option. The TEP optimization problem is run by integrating C^ts from step 5. This optimization problem will be solved much faster than the integrated TEP because it should select fewer new lines when some of them are already selected by the relaxed problem (step 3), and it considers fewer contingencies (size of VCL is much less than the number of total branches N_l). The selected plan by this option is typically near optimal although there is no guarantee for optimality for this option in general, but we quantify the quality of this result by calculating an upper bound for the possible deviation from optimality (see section 2.3.3 for more detail).

- Option (ii): System S_o is considered as the base case system for TEP optimization problem with contingencies. As the result of step 3 are not used as a part of the base case here, this option needs some more simulation time, but it is still much faster than the integrated model as we use a short list of contingencies. Based on Theorem 1 in section 2.3.3, the result of option (ii) is optimal.

The performance of these two options is compared in section 2.4.

Step 7: Run DC-SCOPF with all contingencies for the selected plan in step 6. If there is no violation, for option (i), we have near optimal expansion plan that satisfies $N-1$ contingency analysis and for option (ii) the selected plan will be optimal. (See Theorem 1 in section 2.3.3 for conditions for optimality). Otherwise, update N_u and S_u based on results of step 6, and go back to step 4 to add new possible important lines to VCLs.

In the proposed method, Contingency Identification Index of line l (CII_l) measures the average overload on transmission lines when line l is on outage, and Variable Contingency List (VCL) represents a list of network lines whose single outage causes high overload (more than a predefined threshold) in other lines in the network. The mathematical derivation of CII and VCL are given in section 2.3.2. The performance of the proposed method is also discussed in more detail in section 2.3.3. The flowchart in Figure 2.2 shows important stages of the proposed method (dashed boxes represent the 7 steps).

2.2.4 A Descriptive Example

A simple descriptive example is developed to clarify different steps of the proposed algorithm. Figure 2.3 (a) represents input data for this case

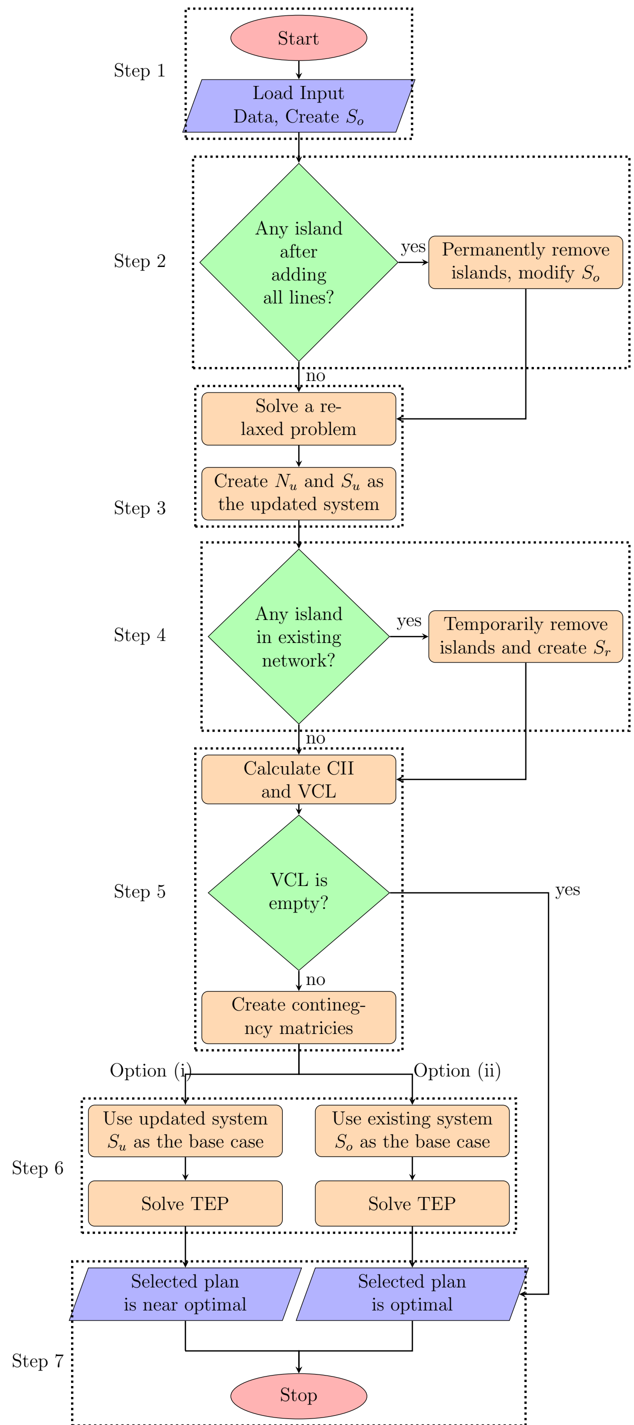


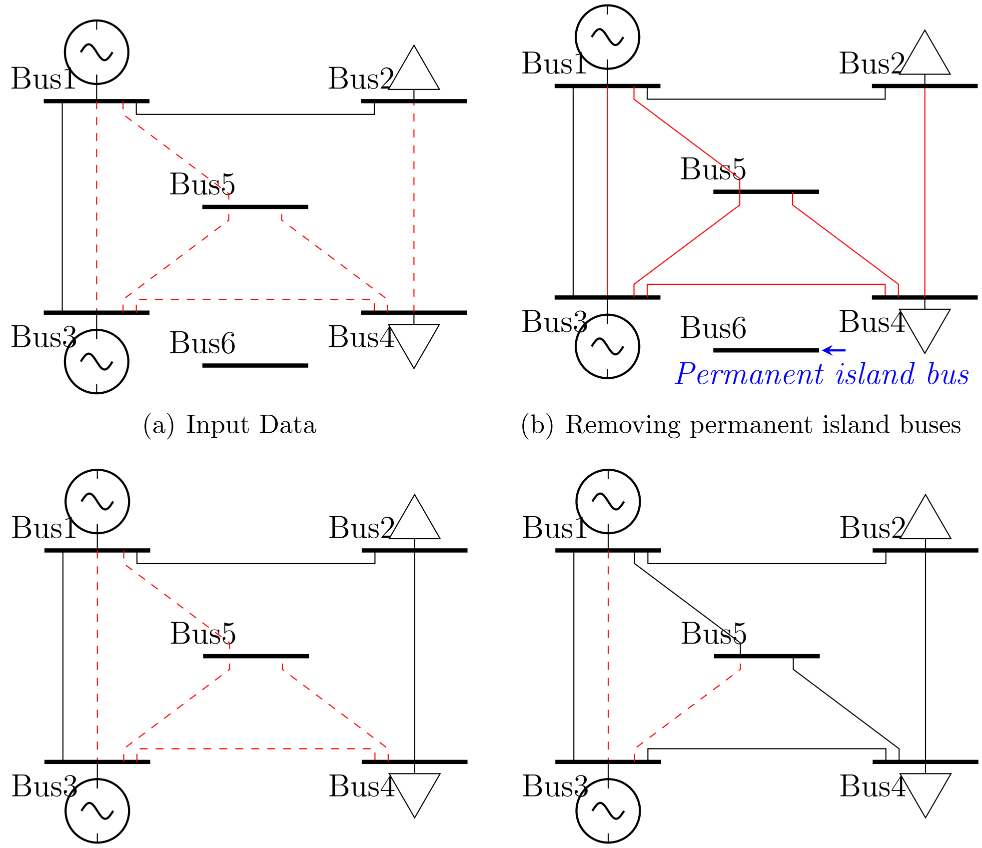
Figure 2.2: Flowchart of the proposed VCL algorithm

study. It contains 6 buses, with 2 generators, 2 loads (load at bus 4 is a new load center that is going to connect to the network), 2 existing (solid black) lines and 5 candidate (dashed red) lines, and the base case S_o contains buses 1 to 6, lines 1-3 and 1-2, two generators and two load centers. It might be argued that buses 5 and 6 can be considered as candidate buses, but as we do not make decision about building/not building a new bus in our model, we represent them as a part of existing network despite them not initially being connected. In step 2, which is shown in Figure 2.3 (b), it is tentatively assumed that all candidate lines are built (shown with solid red lines). In this condition, bus number 6 is an island bus in the system. This may happen because of missing data in the existing/candidate network, a typographical error in the bus name etc., which can happen when working with large scale data. As this bus will never be connected to the network, it is deleted from base case (S_o), and the modified base case is a system with 5 buses, 2 existing lines, 2 generators and 2 load centers. In step 3, as shown in Figure 2.3 (c), the line between bus 2 and 4 is selected by the relaxed problem to be built to supply demand at bus 4, and an updated system S_u is created that includes S_o together with line 2-4. In step 4, we check for islanding again, because for step 5 we need a network without any island. As shown in Figure 2.3 (c), in step 3, candidate lines connected to bus 5 are not selected to be built, so bus 5 is still an island in S_u. It is temporarily removed from S_u to form the reduced system S_r. In step 5, S_r is used to create VCLs. In step 6, there are two different options for solving TEP: option A that uses S_u as the base case and solve TEP with selected contingencies. In this case, the selected line in step 3 (line 2-4) is considered as a part of the base case (S_u) for optimization problem in step 6, therefore there is no guarantee for optimality and the selected plan is near optimal. Lines 1-5, 3-4, 4-5 are added to S_u in this step (shown with solid black lines in Figure 2.3 (d)). Option (ii) uses S_o as the base case for solving TEP with contingency analysis and the selected line in step 3 (line 2-4) is considered as a candidate line again (so the solver makes decision in step 6 to build this line or not). In this case based on Theorem 1, the result of step 6 is optimal. Compared to option (i), option (ii) needs more computational time but still much less than the integrated model. Option (ii) adds lines 1-5, 2-4, 3-4 and 4-5 to S_o for this case (shown with solid black lines in Figure 2.3 (d)).

2.3 Mathematical Formulation and Model Performance Discussion

2.3.1 MIP Formulation with Variable Contingency Matrix

The TEP optimization problem with contingency analysis is formulated as a Mixed-Integer Programming (MIP) problem. The objective function (2.13) contains transmission investment cost plus system operation cost that includes wind and load curtailment penalties and generation production costs for different operation periods. For power system operation cost modeling, security-constrained unit commitment (SCUC) is approximated with individual SCOPF for each load block, because usually a limited number of load blocks is considered to simulate system operation, and as selected load



(a) Input Data

(b) Removing permanent island buses

(c) Results of Step 3

(d) Results of step 6

Figure 2.3: An example for explaining steps of the proposed method

blocks do not necessarily represent sequential operation hours in a day or a week, it is not accurate to integrate them into a SCUC model. Moreover, modeling power system operation using unit commitment adds new binary decision variables that will significantly increase computational time. Recently, authors in [50] developed a tight convex approximation for SCUC that is polynomially solvable, which can be integrated into TEP to tractably represent unit commitment in power system operation. The implicit assumption in our model is that the system has enough ramp rate capability to cope with hourly net load ($Load - Wind$) variations. In this formulation we only integrate contingencies from the VCL algorithm, but in the integrated model for TEP (that we used to compare our results with), single outage of all lines is integrated into TEP optimization problem formulation.

$$Z^{*}{=}\min_{\boldsymbol{x},\boldsymbol{p},\boldsymbol{\theta},\boldsymbol{r},\boldsymbol{f},\boldsymbol{CW}} \sum_{N_n} \zeta_l x_l {+} \sum_{T} \Big[\sum_{N_s^t} (\sum_{N_b} q_i r_{i,c}^t) \Big] + \sum_{T} \Big[\sum_{N_b} \gamma_i CW_i^t {+} \sum_{Ng} Co_g^t p_g^t \Big] \tag{2.13}$$

$$\text{st.} \quad -\sum_{L_k} f_{l,c}^t {+} \sum_{G_k} p_g^t {+} r_{k,c}^t {=} d_k^t, \forall t, k, c \tag{2.14}$$

$$-M_l(1 - C_{l,c}^t x_l) \leq f_{l,c}^t {-} B_{l,l} \Delta\theta_{l,c}^t, \forall t, l, c \tag{2.15}$$

$$M_l(1 - C_{l,c}^t x_l) \geq f_{l,c}^t {-} B_{l,l} \Delta\theta_{l,c}^t, \forall t, l, c \tag{2.16}$$

$$CW_i^t \geq \sum_{W_k} (P_g^{max,t} - p_g^t), \forall t, i \tag{2.17}$$

$$(C_{l,c}^t x_l) f_l^{min} \leq f_{l,c}^t \leq f_l^{max} (C_{l,c}^t x_l), \forall t, l, c \tag{2.18}$$

$$P_g^{min,t} \leq p_g^t \leq P_g^{max,t}, \forall t, g \tag{2.19}$$

$$0 \leq r_{i,c}^t \leq d_i^t, \forall t, i, c \tag{2.20}$$

$$-\frac{\pi}{2} \leq \theta_{i,c}^t \leq \frac{\pi}{2}, \forall t, i, c \tag{2.21}$$

$$CW_i^t \geq 0, \forall t, i \tag{2.22}$$

$$x_l {=} 1, \forall l \in N_o \tag{2.23}$$

$$x_l \in \{0, 1\}, \forall l \in N_l \tag{2.24}$$

In equation (2.13), N_s^t includes normal and single contingency states of the system for each load block t, so the size of N_s^t is equal to $|N_s^t| = 1 + |VCL^t|$. In the objective function, load shedding is penalized under normal and single contingency operation states to prevent load shedding in all N_s^t states in order to satisfy $N-1$ criterion. For wind curtailment and generation dispatch costs, only normal operation condition is considered. Equation (4.6) enforces power balance at each bus. Equations (4.7) and (4.8) show DC representation of flow in transmission lines, the second law of Kirchoff. In these equations, C^t represents contingency matrix for load block t. This matrix contains 0 and 1 as the status of lines (1 for lines in service and 0 for lines on outage). The first column of this matrix represents the normal operating condition, in which no line is on outage, and one line will be on outage in each next column of the matrix based on VCL^t results. The size of this matrix is $|C^t| = |N_l| \times (1 + |VCL^t|)$.

$$C^t = \begin{bmatrix} 1 & 0 & 1 & \cdots & 1 \\ 1 & 1 & 0 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}$$

Equation (4.9) measures the aggregated amount of curtailed wind at each bus. $P_g^{max,t}$ is representing the maximum possible output of wind farm g for each load block t, which is no larger than the nominal installed capacity of that wind farm. Equation (4.10) shows flow in all lines should always be between their capacity limits. Pre-contingency limits are based on continuous rating, so-called Rate A, whereas contingency limits are based on short-term Rate B. We expect that $Rate\,B = Rate\,A(1+\alpha)$, with α on the order of several to many %. When a line is out or is not selected to be built, flow for that line is forced to zero. Equation (4.11) enforces power plants to be dispatched between their minimum and maximum limits. It also shows that pre- and post-contingency dispatch of units are the same. As shown in this equation, maximum and minimum capacity of power plants may change in different load blocks based on their available capacity. It provides more flexibility for applying deterministic changes in the capacity of power plants during different time periods that may happen as a result of expansion, retirement, or scheduled maintenance. Equation (4.12) enforces that load shedding at each bus be greater than or equal to zero and less than or equal to the total load at that bus. Equation (4.13) limits voltage angles at each bus to be between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. Equation (4.14) enforces that wind curtailment cannot be negative. Equation (4.15) sets binary decision variables for existing lines to 1. Equation (4.16) shows that each entry of $\boldsymbol{x}$ is defined as a binary variable to make decision on building a new line ($x = 1$ when a line is built and $x = 0$ when a line is not built).

2.3.2 Variable Contingency List (VCL)

To evaluate the impact of a line outage on post-contingency flows of other lines, DC power flow equations can be used to calculate the sensitivity of lines flow on outage of a line in the network. As mentioned in subsection 2.2.2, there are several different formulations to calculate Line Outage Distribution Factors ($LODF$s) [117], [109] and [44]. In [9], authors showed $PTDF$ is not sensitive to loading conditions and for real networks DC-PTDFs are very close to AC $PTDF$s. By using the direct calculation method proposed in [44], the impact of outage of line l on post-contingency flow of line m i.e. $LODF_{m,l}$ for $l \neq m$ is calculated using the following equations:

$$PTDF_{l,l} = B_{l,l}\Psi_l^T[Y]^{-1}\Psi_l \tag{2.25}$$

$$PTDF_{m,l} = B_{m,m}\Psi_m^T[Y]^{-1}\Psi_l \tag{2.26}$$

$$LODF_{m,l} = PTDF_{m,l}(1 - PTDF_{l,l})^{-1} \tag{2.27}$$

$$f_{m,l} = f_m + LODF_{m,l}f_l \tag{2.28}$$

where Y is the reduced admittance matrix, in which the column and the row related to the reference bus is removed. Ψ is the reduced bus-branch incidence matrix (the row related to the reference bus is removed), and Ψ_l represents the l^{th} column of this matrix that has values 1 and -1 for two ends of line l and 0 for other buses. For lines connected to the reference bus, Ψ_l has only one non-zero element. Equations (2.25)-(2.27) show how to calculate $LODF$ for line m when line l is on outage. If we assume line l connects bus k to bus n, then $PTDF_{m,l}$ shows the impact of injecting 1 MW at bus k and withdrawing 1 MW from bus n on flow in line m (on flow in line l for $PTDF_{l,l}$). As we are dealing with $N-1$ contingency analysis in this chapter, the LODF formulation represents the single contingency case. For multiple contingencies see [44]. Equation (2.28) calculates post-contingency flow in line m. For line

l, post-contingency flow will be zero as line l is out, so $LODF_{m,l}$ is equal to -1 when $m = l$.

To calculate Contingency Identification Index (CII), we first should find lines such that their post-contingency flows violate their capacity limits. Define:

$$\Gamma_{m,l}^{t} = \frac{f_{m,l}^{t} - f_{m}^{max}}{f_{m}^{max}} \qquad ,\forall m, l \in N_u, \forall t \in T \tag{2.29}$$

$$\Phi_{l}^{t} = \{m \in N_u \,|\Gamma_{m,l}^{t} \geq \alpha\} \quad ,\forall t \in T, \forall l \in N_u \tag{2.30}$$

$$CII_{l}^{t} = \begin{cases} \frac{\sum_{m \in \Phi_{l}^{t}} \Gamma_{m,l}^{t}}{|\Phi_{l}^{t}|}, & if\ |\Phi_{l}^{t}| \neq 0 \quad ,\forall t \in T, \forall l \in N_u \\ 0 \qquad\quad , & if\ |\Phi_{l}^{t}| = 0 \quad ,\forall t \in T, \forall l \in N_u \end{cases} \tag{2.31}$$

$$VCL^{t} = \{l \in N_u \,|CII_{l}^{t} \geq \alpha\} \quad ,\forall t \in T \tag{2.32}$$

where $\Gamma_{m,l}^{t}$ in (4.17) evaluates over/under loading in line m compared to Rate A (f_{m}^{max}) when line l is out. Equation (4.18) selects lines such that their overload exceeds emergency capacity (lines with more than α% overload). If no overload is accepted during contingencies then $\alpha = 0$ and $Rate\,B = Rate\,A$. When line l is on outage, equation (2.31) shows how to calculate CII_{l}^{t} for different load blocks. Higher CII for a line means that its outage causes more violation in the rest of the network, so it is a more important line for contingency analysis. Equation (4.20) selects lines that should be included in the contingency matrix and creates VCL^{t} for each load block t.

By using the original $LODF$ formulation (equations (2.25)–(2.27) proposed by [44]), the proposed index would ignore single circuit radial lines from the contingency list. This is because $LODF_{m,l}$ of these lines will be zero for

$m \neq l$ and $LODF_{l,l} = -1$ so the value of CII_l^t is zero for these lines, and they will not be included in VCL^t. But outage of these lines will cause islanding and may result in load shedding or generator outage, which violates the $N-1$ criterion. In some cases, this might be acceptable. However, to modify the index in a way that captures these lines as well, $LODF_{l,l}$ related to these single circuit radial lines is set to a large positive number. This causes large artificial post-contingency flows on these lines and will not have any negative impact on CIIs of other lines. So, these lines will be added to VCL as well, and VCL contains all important lines for contingency analysis.

2.3.3 Model Performance Discussion

This method is developed to make it possible for transmission expansion planners to integrate $N-1$ contingency analysis into systematic planning for large scale power systems. During contingencies, system operators utilize short-term ratings of transmission equipment for a limited time, which is two hours in ERCOT, for example, to prevent possible load shedding and cascading outages. Typical values for α are between 5% to 10% above continuous rating. Networks with tighter capacity may set higher short-term rates, and it is possible for overload limits to vary by lines or even be based on pre-contingency flows [25, 28].

Network loading and short-term rating are two main parameters that affect the size of VCL. Based on network configuration, load level and generation dispatch, the number of lines in VCL may vary. For strong networks in light load condition, VCL might be empty, and for weak networks with high load level the size of VCL will be large. Decreasing short-term rating (decreasing the value of α) will result in increasing the size of VCL and computational time. But this reduction makes the feasible region smaller and the dispatch problem harder to solve, which significantly affects the performance of TEP optimization problem with integrated MIP model as well.

Usually in real power systems, for most loading conditions overloads only occur for outage of a small fraction of lines, and the VCL algorithm finds these lines and include them in TEP instead of considering all lines for contingency analysis or ignoring $N-1$ criterion.

If the optimal answer of a relaxed version (with fewer constraints) of an optimization problem is feasible for the original problem (with all constraints), that answer will be optimal for the original problem based on the following well-known theorem [10].

Theorem 1. Lets $\mathbb{S} \subseteq \overline{\mathbb{S}} \subseteq \mathbb{R}^n$, $f : \mathbb{R}^n \to \mathbb{R}$ and consider the problems: $\min\limits_{\chi \in \mathbb{S}} f(\chi)$, $\min\limits_{\chi \in \overline{\mathbb{S}}} f(\chi)$, and suppose they both have minima and minimizers. Then:

1. $\min\limits_{\chi \in \mathbb{S}} f(\chi) \geq \min\limits_{\chi \in \overline{\mathbb{S}}} f(\chi)$,
2. if $\chi^\star \in \arg\min_{\chi \in \overline{\mathbb{S}}} f(\chi)$ *and* $\chi^\star \in \mathbb{S}$ *then* $\min\limits_{\chi \in \mathbb{S}} f(\chi) = \min\limits_{\chi \in \overline{\mathbb{S}}} f(\chi)$ *and* $\arg\min_{\chi \in \mathbb{S}} f(\chi) = (\arg\min_{\chi \in \overline{\mathbb{S}}} f(\chi)) \cap \mathbb{S}$. ∎

Constraints in an optimization problem form the feasible region, and increasing the number of contingencies may result in a tighter feasible region.

In applying Theorem 1, $\mathbb{S}$ and $\overline{\mathbb{S}}$ represent the feasible region for TEP problem with, respectively, all and only some of (step 6) the contingency constraints. Based on this Theorem, if the optimal answer of TEP in step 6 is feasible for the original problem with all contingencies, the it will in fact also be optimal for the original problem. To check the feasibility, we run DC-SCOPF with all contingencies with fixed binary decision variables (based on the result of step 6), and if there is no violation, it means our answer is in the feasible region of the original problem ($\mathbb{S}$), and it is therefore optimal.

It should be mentioned that Theorem 1 is only applicable for option (ii) in step 6 if the selected plan in this step is feasible for TEP with all contingencies. For option (i) in step 6, the final plan is expected to be near optimal, but we cannot guarantee optimality as some of the selected lines in step 3 may not be selected by option (ii) in step 6. But it is faster than option (ii) as it should select fewer new lines, and it is the planners' choice to select option (i) or B depending on their need. As the result of option (i) is considered an upper bound for the optimal TEP, we can quantify the quality of this result by calculating a lower bound. With no extra computational effort, the maximum error (sub-optimality) can be calculated by considering the result of step 3 as a lower bound answer for TEP and using equation (4.1):

$$Maximum\,Error = \frac{Upper Bound - Lower Bound}{Lower Bound} \tag{2.33}$$

It is also possible to use Branch and Bound (BB) algorithms [56] to get a better lower bound and improve the gap between upper and lower answers. Details of BB is not in the scope of this dissertation. In section 2.4, the impact of using option (i) or (ii) in step 6 on both optimal plan and computational time are evaluated.

The solver's optimality gap is another parameter affecting computational time. Solvers use a predefined maximum optimality gap as a stopping criteria when they solve MIP problems (the default value of this gap is 0.01% in GUROBI and CPLEX). Adjusting this gap directly affects the simulation time for both TEP in step 6 and the integrated model. As this impact is not linear, the relative performance of the proposed method may change by changing the solver's optimality gap. Increasing the optimality gap will decrease simulation time, but it may affect the selected plan as well. This issue is investigated in section 2.4.

2.4 Case study and Simulation results

All illustrated results in this section have been obtained from a personal computer with 2.0-GHz CPU using MATLAB R2014a [74] and YALMIP R20140221 package [60] as a modeling software and GUROBI 5.6 [45] as the solver. Two different case studies consisting of 13-bus system and reduced ERCOT network with 317-buses are considered. The MATLAB built-in function `tic toc` is used to evaluate elapsed wall clock time ("Total Time"). For each case study, we solve the TEP problem for both options (i) and (ii) and the integrated model (in which all contingencies are integrated into TEP) to compare the quality of results and computational time.

2.4.1 13-Bus System

This 13-bus system is a simplified version of the ERCOT network that is developed for educational purposes (see Figure 2.4). This case study has 13 buses, 33 branches, 16 power plants, and 9 load centers. Bus No.13 is a new load center that is going to be connected to the network in the planning time horizon, so it will be an island bus in our base case. Two different loading and wind production blocks are considered to represent the whole year (50% of year for each load block). The number of candidate lines is 36, which represents potential expansion and reinforcement in the whole area. Load, generation, existing and candidate lines data are given in Table 2.1–2.3. Two load/wind blocks in Table 2.1 are considered for $t = 1, 2$. We considered transmission investment as having on overnight cost of \$1M/mile (16% of this cost is considered as annual investment cost). Value Of Loss Load (VOLL) is \$9000/MWh (based on ERCOT market price cap [26]) and wind curtailment penalty is set to \$500/MWh, and solver's optimality gap is set to 0.1%. A high penalty cost for wind is assigned to force TEP to reinforce the network to transfer all wind output from west Texas to central and east Texas.

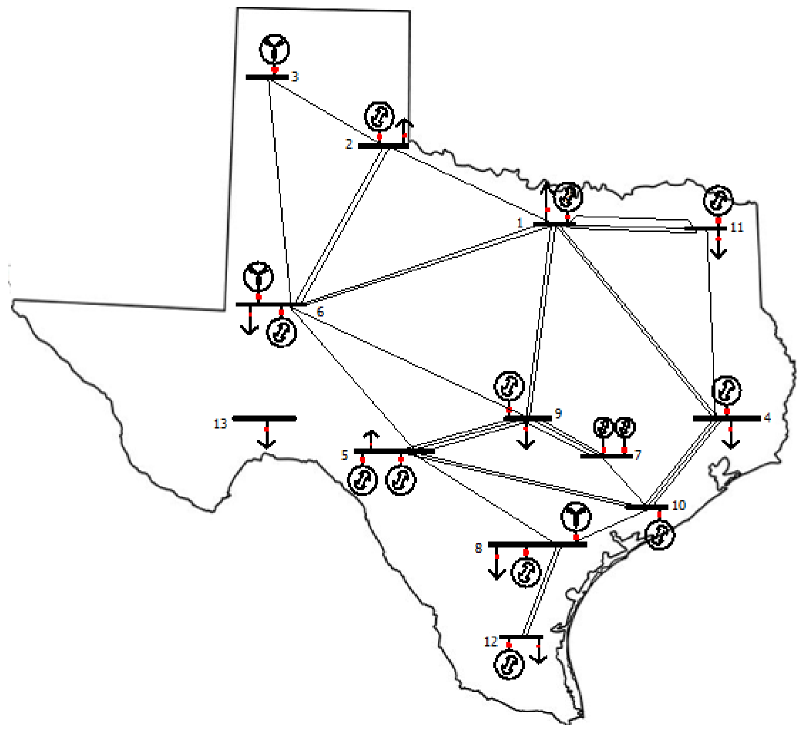


Figure 2.4: 13 bus network with existing lines, generators and loads.

Table 2.1: Load and Generation Data in [MW]

Bus	Gen	Load1	Load2	Wind1	Wind2
1	21374	22964	19519	0	0
2	2811	475	403	0	0
3	0	0	0	3000	3500
4	24292	24582	20895	0	0
5	8233	5960	5066	0	0
6	6216	5305	4509	4000	5500
7	1208	0	0	0	0
8	5881	4417	3755	1000	3200
9	4657	8383	7125	0	0
10	2750	0	0	0	0
11	3262	547	465	0	0
12	2503	3367	2862	0	0
13	0	1000	850	0	0

Table 2.2: Existing Transmission Network Data

From	To	Susceptance [P.U.]	Capacity [MW]
2	1	13.89	1000
1	4	8.20	625
1	4	8.20	625
1	6	8.85	812.5
6	1	8.85	912.5
1	9	11.11	875
1	9	11.11	937.5
1	11	15.87	1125
1	11	15.87	1125
1	11	15.87	1125
3	2	13.33	1062.5
2	6	12.35	1125
6	2	12.35	1125
3	6	9.26	875
4	10	27.78	1125
4	10	27.78	1125
4	10	27.78	1125
11	4	9.62	1000
6	5	8.55	937.5
8	5	15.87	812.5
9	5	25.00	1750
9	5	25.00	1750
5	9	25.00	1750
5	10	12.35	875
5	10	12.35	812.5
6	9	8.55	875
9	7	34.48	1250
9	7	34.48	1250
9	7	34.48	1250
7	10	22.22	1750
8	10	16.95	875
8	12	37.04	1312.5
8	12	37.04	1312.5

Table 2.3: Candidate Lines

From	To	Susceptance [P.U.]	Capacity [MW]	Length [mile]
2	1	13.89	1000	144
2	1	13.89	1000	144
1	4	8.20	625	243
1	4	8.20	625	243
1	6	8.85	812.5	225
6	1	8.85	812.5	225
1	11	15.87	1125	126
3	2	13.33	1062.5	150
3	2	13.33	1062.5	150
2	6	12.35	1125	162
6	2	12.35	1125	162
3	6	9.26	875	216
3	6	9.26	875	216
4	10	27.78	1125	72
11	4	9.62	1000	207
6	5	8.55	937.5	234
6	5	8.55	937.5	234
8	5	15.87	812.5	126
9	5	25.00	1750	81
9	5	25.00	1750	81
6	9	8.55	875	234
6	9	8.55	875	234
7	10	22.22	1750	90
8	10	16.95	875	117
8	10	16.95	875	117
8	12	37.04	1312.5	108
8	12	37.04	1312.5	108
13	6	13.00	1125	173
13	5	20.05	1125	112.2
13	9	10.80	875	208.3
13	6	13.00	1125	173
13	5	20.05	1125	112.2
13	9	10.80	875	208.3
13	6	13.00	1125	173
13	5	20.05	1125	112.2
13	9	10.80	875	208.3

For this case study, first we evaluate the impact of considering $N-1$ contingency analysis on transmission planning compared to ignoring it. In Table 2.4, results for two cases of TEP without and with contingency are shown. In the case with contingency, it will turn out that we need to build three extra lines. By enforcing $N-1$ criterion during operation for the network that was built without considering contingencies (by running DC-SCOPF), there are 938.4 MWh load shedding and 454.4 MWh wind curtailment on average during normal operation conditions. It will result in \$75973.73 M penalty costs per year, which is 810 times more than the difference in investment cost between the two cases of enforcing and not enforcing $N-1$ security in the planning of the network (\$94 M). This example shows that ignoring contingency analysis in planning stage may result in huge operation costs and load shedding, which is against $N-1$ criterion. It should be emphasized that we did not consider demand response and other real-time corrective actions that a system operator may take during an outage. We also considered only two load/wind blocks to represent the whole year; therefore this ratio (810 times) may change with more accurate operation cost modeling.

To show how our proposed method works, we summarize all steps explained in section 2.2.3 for this case. Step 1: load input data: input data is loaded and S_o is created; Step 2: check for islanding: no island; Step 3: solve relaxed problem: Four selected lines in this step are shown in the second column of Table 2.4, N_u and S_u are created; Step 4: islanding check: no island, S_r is created; Step 5: create VCL: VCL is created for two load blocks. α is

Table 2.4: Transmission Expansion Planning with and without considering Contingencies, 13-Bus System

	Without Contingency	With Contingency
Selected Lines	3-2	3-2
	3-6	3-6
	7-10	3-6
	13-6	6-9
	-	7-10
	-	13-6
	-	13-5
Total Cost($ M)	6047.23	6137.83
Investment Cost($ M)	104.83	198.53
Operation Cost($ M)	5942.4	5939.3

set to 0.05, which means during a contingency the remaining lines can tolerate 5% above their continuous rating. VCLs for two load blocks are shown in Table 2.5. The number of selected lines for contingency analysis is different for these two blocks i.e 18 lines for the first load block and 17 for the second one, and both are much less than the 69 lines ($|N_l|$) that should be considered in the integrated TEP optimization problem. Moreover, Table 2.5 shows that a fixed list of contingencies may not be sufficient for satisfying $N-1$ criterion for different operation conditions. As VCL is not empty, we should create contingency matrices C^1 and C^2 for $t = 1$ and 2 respectively; Step 6: solve TEP: TEP optimization problem is solved for C^1 and C^2 matrices for both options (i) and (ii). Results of this step will be discussed later in this section. Go to step 7. Run DC-SCOPF with all contingencies, no violation. So, the selected plan is optimal/near-optimal and satisfies $N-1$ criterion.

Table 2.5: VCL for 13-Bus System

	Selected Lines for Contingency Analysis
VCL^1	2 - 1 , 1 - 9 , 1 - 9 , 3 - 2 , 2 - 6 , 6 - 2 , 3 - 6 , 11 - 4 , 6 - 5 , 9 - 5 , 9 - 5 , 5 - 9 , 6 - 9 , 7 - 10 , 3 - 2 , 3 - 6 , 7 - 10 , 13 - 6
VCL^2	2 - 1 , 1 - 6 , 6 - 1 , 1 - 9 , 1 - 9 , 3 - 2 , 3 - 6 , 6 - 5 , 5 - 10 , 5 - 10 , 6 - 9 , 7 - 10 , 8 - 10 , 3 - 2 , 3 - 6 , 7 - 10 , 13 - 6

The integrated TEP with all contingencies is also solved to compare its result and simulation run time with our proposed method. Final results are shown in Table 2.6. The second row shows selected lines by option (i), B, and the integrated model. For this case, all three methods select the same (optimal) plan with the same costs. The last row of this Table shows the total time for simulation. Option (i) is around 2 times faster than option (ii) and more than 75 times faster than the integrated model (this ratio for option (ii) is 33.73) because structural constraints of TEP problem are reduced more than 72% by using VCL algorithm. Based on equation (4.1), the maximum error bound for option (i) is less than 1.5% and for this case the actual gap is 0%.

2.4.2 Reduced ERCOT System

A reduced model of the ERCOT system is provided in [94]. This network contains 317 buses, 427 branches, 489 conventional power plants, 36 wind farms and 182 load centers. The purpose of developing this case was to evaluate the impact of large penetration of wind in the Competitive Renewable Energy Zone (CREZ) area of the ERCOT market and the transmission expan-

Table 2.6: Transmission Expansion Planning for 13-Bus System

	Option (i)	Option (ii)	Integrated model
Selected Lines	3-2 3-6 3-6 6-9 7-10 13-6 13-5	3-2 3-6 3-6 6-9 7-10 13-6 13-5	3-2 3-6 3-6 6-9 7-10 13-6 13-5
Total Cost($ M)	6137.83	6137.83	6137.83
Investment Cost($ M)	198.53	198.53	198.53
Operation Cost($ M)	5939.3	5939.3	5939.3
Total Time (sec)	3.91	8.2	276.6

sion requirements to transfer wind power to central and east Texas. For this reason, west Texas is simulated in detail, and the rest of the ERCOT area is aggregated to three zones as delivery points of CREZ. All costs and related parameters are set the same as the 13-bus system.

It is assumed that all network lines can be reinforced, which means the number of candidate lines (and corresponding binary variables) is 427 (compared to 5 and 17 candidate lines for IEEE-118 Bus system used by [54] and [122] respectively). The number of binary variables in a MIP optimization problem has a huge impact on computation time (because of the combinatorial nature of the problem), and the large number of binary decision variables makes this case study a very challenging problem. Solving the relaxed version of the problem (step 3) adds 4 new lines that will decrease computational time for option (i) in step 6. VCL for the reduced ERCOT system is shown in Table 2.7. In this network with 433 existing lines ($|N_u|$) and 421 candidate lines, only 7 lines are selected for contingency analysis for this loading condition. Outage of these lines will activate reliability constraints in the TEP optimization problem and will affect the feasible region and the optimal answer. As observed earlier, most of the time in real networks, there are only a few lines whose outage may cause overload on other lines, and we mostly do not need to consider all lines in contingency analysis to get an optimal plan that satisfies the $N-1$ criterion. This list of lines may vary by changing network conditions, and CII^t recognizes these important lines for each load block t.

The selected optimal expansion plan for this network is summarized in Table 2.8. The second column shows the selected network when we ignore contingency analysis in TEP, as obtained from step 3. Ignoring contingency analysis, we need to build 4 new lines with \$25.033 M investment cost. If we also ignore $N-1$ criterion in dispatching power plants in the operation stage (running OPF), the operation costs will be \$43618 M as stated in Table 2.8. But if the system operator should satisfy $N-1$ criterion for operation (running SCOPF instead of OPF), the operation costs of this network will be \$104170 M as a result of load shedding and wind curtailment penalty costs (during normal operating condition), which is around 2.4 times the operation costs for the case without contingency analysis (using OPF).

Considering contingency analysis in TEP added 5 new lines (columns 3 and 4 in Table 2.8 for options (i) and (ii) to the plan shown in column 2 that

Table 2.7: VCL for Reduced ERCOT System

	Selected Lines for Contingency Analysis
VCL	1065 - 1064 , 1066 - 1065 , 5905 - 5902 , 46220 - 7270 , 7670 - 7668 , 5905 - 5902 , 7670 - 7668

was selected without considering contingencies. Investment costs increased by more than 100% compared to ignoring contingency analysi. This is \$26.7 M extra investment cost, which is 0.044% of the extra operation cost that would result from ignoring contingency analysis during planning but enforcing it during operation. This result clearly demonstrates the effectiveness of VCL algorithm in selecting important lines for contingency analysis and their impact on the final expansion plan. This plan satisfies $N-1$ criterion and its operational cost is less than 42% of the previous case (\$60530 M saving). Columns 3 and 4 show the same lines are selected by options (i) and (ii) in step 6, but option (i) is around 5.7 times faster. As shown in column 5, results could not be obtained (still more than 14% optimality gap) for the integrated model even after 10 days (compared to 407.8 seconds with the proposed method). To verify the selected plan satisfies $N-1$ criterion for all contingencies, a full DC-SCOPF is run. No load shedding or wind curtailment shows the selected plan satisfies $N-1$ criterion, and it has been obtained more than 3500 times faster than the integrated model.

As stated before, the solver's optimality gap for MIP is set to 0.1% for numerical analysis in this chapter. This value directly affects the computational time and decreasing this gap can be expected to increase simulation

Table 2.8: Transmission Expansion Planning for Reduced ERCOT System

	Step 3	Option (i)	Option (ii)	Integrated model
Selected Lines	5905 - 5902 7670 - 7668 90000 - 42500 90002 - 5915 - - - - -	1065 - 1064 1066 - 1065 5904 - 5902 5905 - 5902 7670 - 7668 13430 - 3430 90000 - 42500 90001 - 5905 90002 - 5915	1065 - 1064 1066 - 1065 5904 - 5902 5905 - 5902 7670 - 7668 13430 - 3430 90000 - 42500 90001 - 5905 90002 - 5915	No feasible solution is found
Total Cost($ M)	43643	43696	43696	-
Investment Cost($ M)	25.033	56.725	56.725	-
Operation Cost($ M)	43618	43640	43640	-
Total Time (sec)	243.65	407.8	2325.27	10+ days

time exponentially. In Table 2.9, the case study is run for different optimality gaps i.e 0.01%, 0.1% and 3% for option (i). The last row in this table shows the computational time that the solver (GUROBI in this case) needs to solve the MIP problem for different optimality gaps. Computational time for 0.01% optimality gap is 155 times more than 0.1% gap, and this ratio is 1606 for 3% gap that demonstrate the significant impact of optimality gap on simulation time. The selected plans for different gaps are shown in the second row of Table 2.9. By adjusting optimality gap to 3% for this case study, a different plan is selected compared to 0.01% and 0.1% gaps. It shows that changing optimality gap not only affects the simulation time but also may change the selected plan; therefore there is a trade-off between computational time and optimality gap selection. However, in practice and for large scale networks, getting to tight optimality gaps like 0.1% or 0.01% is extremely computationally expensive, and we usually should accept optimality gap between 1% to 5% to be able to get an answer in a reasonable time.

2.5 Summary

The impact of contingency analysis on transmission capacity expansion planning is evaluated. Simulation results show that ignoring contingency analysis in TEP may cause load shedding and huge extra operation costs when system operators should satisfy $N-1$ criterion. In most loading conditions, it is not necessary to consider all lines in contingency analysis, because the single outage of some lines will not cause overload on other lines. Since the constraints related to these lines will remain passive in the optimization problem, we proposed a systematic and effective method to integrate necessary contingencies into TEP such that the final plan satisfies the $N-1$ criterion. The proposed method is organized in 7 steps, by solving the relaxed version of problem (TEP without contingency analysis) in step 3, and adding important contingencies as additional constraints to the TCEP optimization problem in step 6. Contingency identification index is developed and integrated to the process to detect important lines for contingency analysis and creates variable contingency lists (VCLs) for different network configuration and loading conditions. Step 6 has two options for solving TCEP with contingencies i.e. option (i) that provides an upper bound (with known optimality gap) for TEP with less computational effort and option (ii) that provides the optimal answer, and depending on the whole planning process option (i) or (ii) may be

Table 2.9: Evaluating the impact of optimality gap on optimal answer and simulation time

	0.01% gap	0.1% gap	3% gap
	1065 - 1064	1065 - 1064	1066 - 1065
	1066 - 1065	1066 - 1065	5904 - 5902
	5905 - 5902	5905 - 5902	7670 - 7668
	5904 - 5902	5904 - 5902	13430 - 3430
	7670 - 7668	7670 - 7668	46220 - 7270
	13430 - 3430	13430 - 3430	90000 - 42500
	90000 - 42500	90000 - 42500	90000- 44000
	90001 - 5905	90001 - 5905	90000 - 44500
	90002 - 5915	90002 - 5915	90000 - 45972
Selected Lines	-	-	90000 - 46220
	-	-	90000 - 46500
	-	-	90001 - 1967
	-	-	90001 - 3390
	-	-	90001 - 3391
	-	-	90001 - 13430
	-	-	90001 - 40600
	-	-	90002 - 3430
	-	-	90002 - 5915
	-	-	90002 - 7270
Total Cost($ M)	43696	43696	44109
Solver Time (sec)	8435.3	54.28	5.25

selected by planners. The method is implemented on a reduced ERCOT network with 317 buses and 427 binary decision variables that makes this MIP problem very hard to solve. The results show that using the relaxed problem and effective selection of lines for contingency analysis will significantly reduce computational time (more than 3500 times for this case study), and make it practically possible to integrate contingency analysis into systematic TEP optimization problem for large scale power systems with several load blocks.

Chapter 3

Reducing Candidate Lines List for stochastic TEP[1]

3.1 Introduction

In recent years, building new transmission lines becomes more and more difficult because of environmental concerns, and it can take several years to plan and build new transmission lines, raising the need for long term transmission planning (10+ years). Increasing the planning time horizon will result in more uncertainty in future generation and load capacity/locations that usually are distributed over a wide geographical area, resulting in a large candidate lines list (CLL) in early stages of transmission planning especially when transmission routing is used to investigate different alternative paths to connect two buses.

As stated before, TEP is a large scale, non-convex and nonlinear optimization problem, and the number of candidate lines (which is equal to the number of binary variables in the optimization problem) significantly affects

[1]Mohammad Majidi-Qadikolai and Ross Baldick. Reducing the number of candidate lines for high level transmission capacity expansion planning under uncertainties. In North American Power Symposium (NAPS), 2015, pages 1-6, Oct 2015. Authors had equal contributions.

the computational time for TEP. A shorter CLL decreases the problem size (combinatorial nature of integer programming) and makes it possible to solve larger scale network with more accurate modeling (for example developing more scenarios to capture uncertainties). In practice, usually expert knowledge is used to remove some candidate lines from CLL. But in long-term planning with uncertainties in future generation and load locations and intermittent resources (solving stochastic TEP instead of deterministic TEP, which is solved in Chapter 2), it will be much harder to detect unimportant candidate lines based on expert knowledge. In this chapter, a heuristic method is developed that removes some lines from CLL in a systematic way by considering their impact on alleviating existing congestion in the base case.

3.1.1 The Basic Idea

In contingency analysis in power systems, it is common to limit monitored lines list based on geographical location and loading condition, and there are several heuristic methods for this purpose [117]. Authors in [120] suggested that limiting monitored lines for contingency analysis to the lines close to the line on outage will reduce simulation time for TEP. They divided the IEEE-118 bus system into three zones (see Figure 3.1), and argued that, for a line outage in zone 1, it is sufficient to monitor lines in this zone, instead of all lines in the network.

It can be observed that LODFs will be close to zero for lines far away from the line on outage. Therefore, flow on those lines will not be affected strongly as a result of an outage far away from them. So, their flow would not typically increase above emergency limits subsequent to a far away contingency.

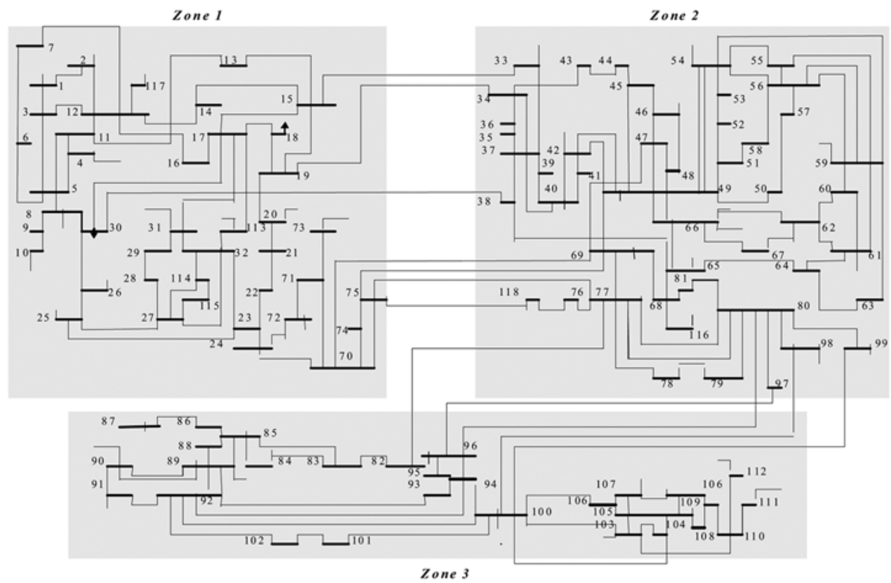


Figure 3.1: IEEE 118-bus system

This is also true for closing a line. Building a new line can be interpreted as closing an opened line. It is expected that building a new line will have more impact on lines on its nearby rather than on lines far away from it. Line Closure Distribution Factor (LCDF) can be used for such sensitivity analysis. The detail related to LCDF is given in subsection 3.3.2. It is also possible to limit monitored lines to those that are heavily loaded (in the base case) because we are expecting to add new lines that decrease overload/congestion in the base network. We aggregated these two heuristics into our model to reduce computational time in an automated and tractable way. The model performance is discussed in subsection 3.3.3.

The proposed method is useful for large scale problems in which reducing computational burden is critical. It is applied to a reduced ERCOT network with 317 buses and 427 candidate lines for different number of scenarios, and the result shows 77%-89% reduction in CLL size depending on the number of scenarios, and we get the results several hundred times faster compared to the case that there is no reduction in CLL.

The rest of this chapter is organized as follows: in section 3.2 the proposed method is explained, and it is followed by the mathematical formulation of stochastic transmission expansion planning and a discussion of model performance in section 3.3. In section 3.4, the proposed method is implemented on our ERCOT case study, and results are compared with the original method. A summary is given in section 3.5.

3.2 The Proposed Candidate Line Reduction Method

In the early stage of long-term planning, planners develop a relatively large number of possible candidate lines to cover all areas in the network that may need expansion, and different techniques such as expert knowledge, environmental, and technical analysis are used to reduce this list in the next steps of planning. Using expert knowledge for complicated systems is not usually tractable and depends on individuals' expertise. Evaluating environmental concerns is costly and takes time. The proposed method in this section can be used to reduce the number of candidate lines in the early stage of planning and makes the future steps easier from both the computational perspective and the process costs. The proposed method can be summarized in the following steps:

Step 1: Load Input Data.

Load, Generation, current and candidate transmission components etc are included in input data. The base case system that contains existing lines, load, buses and generators is referred to as S_o. The initial candidate lines list in this step is referred as CLL_o.

Step 2: Temporarily remove island buses from S_o.

In this step, if there is any island bus in the base case it will be temporarily removed from the S_o to form a new reduced system which is referred to as S_r because we need a system without any island in the next steps. It should be noted that candidate lines connected to these island buses also will be removed from CLL_o temporarily to create CLL_r. If a load or generator is connected to any of island buses, set $Flag = 1$, otherwise, set $Flag = 0$.

Step 3: Solve a relaxed version of OPF problem.

In this step, a relaxed version of OPF is run for S_r, in which constraints related to line capacity limits are ignored. By solving this OPF, we can find the lines in the base case system that will be overloaded for the target planning year. If there is no overload in lines, no load shedding or wind curtailment **and** $Flag = 0$, it means our current network can supply the load and we do not need to add any new lines, so go to step 7, otherwise go to step 4.

Step 4: Create monitored lines list (MLL).

In this step, lines with flows more than $\eta\%$ (50%, for example) of their nominal capacity will be added to Monitored Lines List (MLL). The OPF result from step 3 is used to calculate flow violations in existing lines. MLL will be used in the next step for evaluating the impact of adding candidate lines.

Step 5: Update candidate lines list (CLL_u).

Line Closure Distribution Factor (LCDF) is calculated for candidate lines in CLL_r to evaluate the impact of closing each candidate line on the lines in MLL. Candidate lines such that their closure will not decrease flow in at least one line in MLL will be added to Extra Candidate Lines List (CLL_E). Connecting these lines may decrease flow in other lines, but it will not alleviate any congestion in the network as affected lines are already lightly loaded (less than $\eta\%$). Lines in CLL_E will be removed from CLL_o to create an updated candidate lines list, which is referred as CLL_u ($CLL_u = CLL_o \setminus CLL_E$). It should be mentioned that as lines in CLL_E are removed from CLL_o (not CLL_r), all temporarily removed lines in step 2 are included in CLL_u. Mathematical formulation and more discussion about the model are given in sections 3.3.2 and 3.3.3 respectively.

Step 6: Solve TEP optimization problem.

System S_o is used as the base case system and CLL_u as the candidate lines list for this step. TEP optimization problem is run. This optimization problem can be expected to solve faster than TEP with CLL_o because it has less candidate lines (binary variables). Compared to Chapter 2 TEP formulation, operation cost is extended to expected operation cost that integrates uncertainties in future power system operation using different scenarios.

Step 7: The selected expansion plan is expected to be near optimal.

The flowchart in Figure 3.2 shows important stages of the proposed method (dashed boxes represent 7 steps). The performance of the proposed method is discussed in more detail in subsection 3.3.3.

3.3 Mathematical Formulation and Discussion

3.3.1 MIP Formulation for stochastic TEP optimization

Stochastic TEP is a two-stage recourse allocation problem that can be represented in extensive form by approximating the expected operation costs with weighted sum of operation cost for different scenarios. By using DC representation of power flow equations, stochastic TEP is formulated as a MIP problem. The objective function (3.1) contains transmission investment cost and expected operation cost that includes the weighted sum of wind and load curtailment penalties along with generation production costs. Scenarios can be

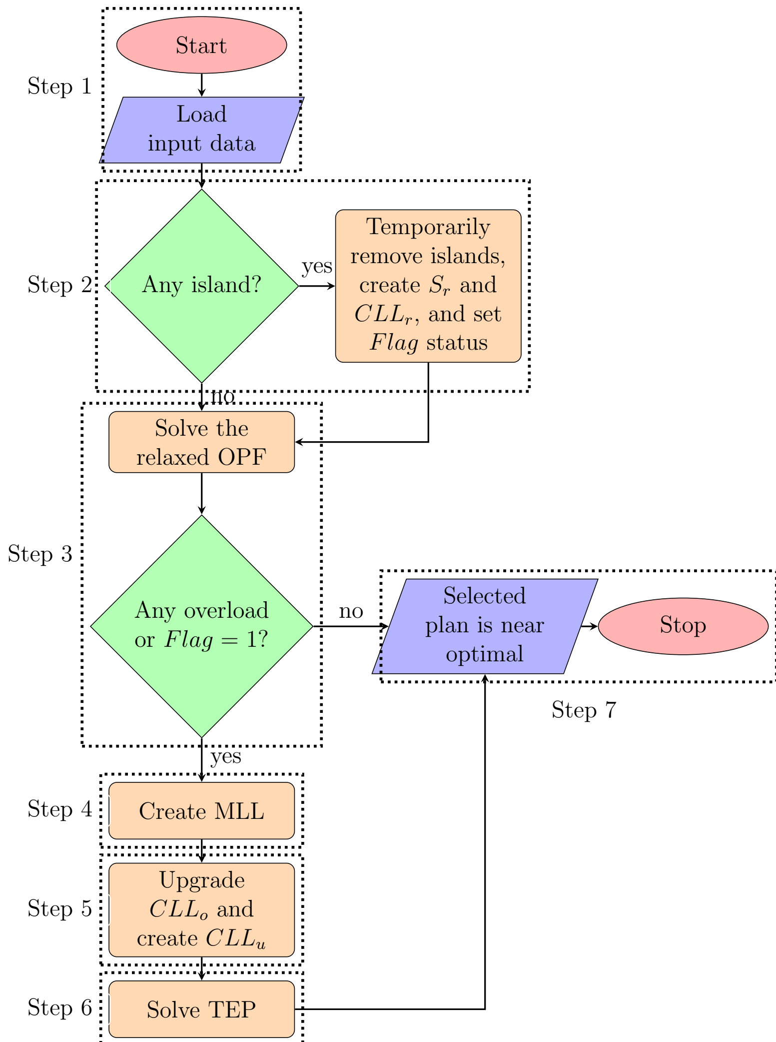


Figure 3.2: Flowchart of the proposed CLL reduction method

generated to model different uncertainties, and in this chapter they represent wind and load uncertainties. The following formulation is very similar to (2.13)–(4.16) in section 2.3.1, except that in (2.13)–(4.16) deterministic TEP with contingency analysis is formulated whereas in (3.1)–(3.12), stochastic TEP with updated candidate lines list and without reliability constraints is formulated.

$$Z^{*}=\min_{\boldsymbol{x},\boldsymbol{p},\boldsymbol{\theta},\boldsymbol{r},\boldsymbol{f},\boldsymbol{CW}} \sum_{CLL_u} \zeta_l x_l + \sum_{\Omega} P^{\omega} \Big[\sum_{N_b} (\gamma_i CW_i^{\omega} + q_i r_i^{\omega}) + \sum_{Ng} Co_g^{\omega} p_g^{\omega} \Big] \tag{3.1}$$

$$\text{st.} \quad -\sum_{L_k} f_l^{\omega} + \sum_{G_k} p_g^{\omega} + r_k^{\omega} = d_k^{\omega}, \forall \omega, k \tag{3.2}$$

$$-M_l(1-x_l) \leq f_l^{\omega} - B_{l,l} \Delta\theta_l^{\omega}, \forall \omega, l \tag{3.3}$$

$$M_l(1-x_l) \geq f_l^{\omega} - B_{l,l} \Delta\theta_l^{\omega}, \forall \omega, l \tag{3.4}$$

$$CW_i^{\omega} \geq \sum_{W_k} (P_g^{max,\omega} - p_g^{\omega}), \forall \omega, i \tag{3.5}$$

$$x_l\, f_l^{min} \leq f_l^{\omega} \leq f_l^{max} x_l, \forall \omega, l \tag{3.6}$$

$$P_g^{min,\omega} \leq p_g^{\omega} \leq P_g^{max,\omega}, \forall \omega, g \tag{3.7}$$

$$0 \leq r_i^{\omega} \leq d_i^{\omega}, \forall \omega, i \tag{3.8}$$

$$-\frac{\pi}{2} \leq \theta_i^{\omega} \leq \frac{\pi}{2}, \forall \omega, i \tag{3.9}$$

$$CW_i^{\omega} \geq 0, \forall \omega, i \tag{3.10}$$

$$x_l = 1, \forall l \in N_o \tag{3.11}$$

$$x_l \in \{0,1\}, \forall l \in N_l \tag{3.12}$$

Equation (3.2) enforces power balance at each bus under each scenario.

Equations (3.3) and (3.4) show power flow in transmission lines, the second law of Kirchoff for all existing and candidate lines under different scenarios. Equation (3.5) measures the amount of curtailed wind at each bus. $P_{W_i}^{max,\omega}$ is representing the maximum possible output of wind farm for scenario ω, which is no larger than the nominal installed capacity of wind farm. Equation (3.6) shows flow in all lines should always be between their capacity limits. When a line is not selected to be built, flow for that line is forced to zero. Equation (3.7) enforces power plants to be dispatched between their minimum and maximum limits. Equation (3.8) enforces that load shedding at each bus be greater than or equal to zero and less than or equal to the total load at that bus. Equation (3.9) limits voltage angles at each bus to be between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. Equation (3.10) enforces that wind curtailment cannot be negative. Equation (3.11) sets binary decision variables for existing lines to 1. Equation (3.12) shows the x is defined as a binary variable to make decision on building a new line ($x = 1$ when a line is built and $x = 0$ when a line is not built).

3.3.2 Updating Candidate Lines List (CLL_u)

Supplying demand and increasing the reliability of the system in an economic way are two main reasons for TEP. To find unimportant candidate lines for expansion so that they can be removed from consideration in order to reduce computational effort, we need to know the impact of building new lines on the flows on congested lines in the network. First we need to find lines that may be congested in the existing network. By running an OPF in which constraints related to line capacities (equation (3.6)) are relaxed, flows in existing lines are calculated and monitored lines list (MLL) is formed from the following equations:

$$Dev_l^{\omega} = f_l^{\omega} - \eta \times f_l^{max}, \forall l, \omega \tag{3.13}$$

$$MLL = \{l \in N_o \,|\, Dev_l^{\omega} > 0, \forall \omega\} \tag{3.14}$$

where (3.13) calculates over/under flow in lines compared to $\eta\%$ of their nominal capacities, and (3.14) adds lines with over flow (more than $\eta\%$ loading) into MLL. f_l^{ω} represents the magnitude of flow in line l under scenario ω (calculated in step 3.)

Line Closure Distribution Factor (LCDF) is used to evaluate the impact of a line closure on lines in MLL.

$$PTDF_{l,(nr)} = (\chi_{ni} - \chi_{ki})\, B_{l,l} \tag{3.15}$$

$$LCDF_{m,l} = \frac{PTDF_{m,(nr)}^{IN} - PTDF_{m,(nr)}^{OUT}}{PTDF_{l,(nr)}^{OUT}} \tag{3.16}$$

$$CLL_E = \{l \in CLL_r \,|LCDF_{m,l} \geq \beta, \forall m \in MLL\} \tag{3.17}$$

$$CLL_u = CLL_o \setminus CLL_E \tag{3.18}$$

$$-1 \leq \beta \leq 0 \tag{3.19}$$

Equations (3.15)-(3.16) show how to calculate LCDF for line m when line l is closed based on [109]. Line l is between bus n and k. χ_{ni} represents element ni of the reduced Z-bus matrix χ. Superscript IN refers to the case that line l is closed, and OUT refers to the base case without line l. Equation (3.17) extracts unimportant lines from reduced candidate lines list CLL_r. β is a parameter that is set by the planner based on required relaxation. Values closer to -1 result in more relaxation and a smaller CLL_u list. Equation (3.18) creates updated candidate lines list (CLL_u) by removing selected lines in (3.17) from original candidate lines list (CLL_o). The original candidate lines list (CLL_o) is an input in our model, and it can be created by planning experts manually or in an automated way.

3.3.3 Model Performance Discussion

This method is developed to reduce the size of candidate lines list (CLL) in the early stages of TEP for large scale power systems with a large CLL_o. As mentioned in 3.1.1, adding a new line mostly affects the lines close to it rather than those that are far away from it in large scale networks, and this characteristic makes the proposed method more effective for large networks (compared to small networks). For small scale networks, running this algorithm will not significantly reduce the size of CLL_o and may negatively affect the selected plan, so it will not be effective. By selecting highly loaded lines for monitoring (MLL), we limit evaluating the impact of adding a new line to the lines with more than $\eta\%$ loading (this loading percent can be set by the planner). The reason for ignoring lines with less than $\eta\%$ loading is that these lines are already lightly loaded and there is no reason to build a new line to further decrease flow on these lines. By setting β close to zero, candidate lines that will not have positive impact on flow (reducing congestion) on monitored lines close to them will be removed from the CLL. If building a line increases flow on all monitored lines close to it, it means this line will have a large negative impact on those lines, so it will not be an appropriate candidate line to be built.

In equation (3.13), multiplying the nominal capacity of lines by η means that if flow in a line is more than $\eta\%$ of its nominal capacity it will be considered as an important line for monitoring in our model. Increasing this value decreases the size of MLL, and limits the number of monitored lines that may reduce the size of CLL_u. Planners can control the size of CLL_u by changing this value.

In equation (3.17), lines are selected that have $LCDF_{m,l}$ greater than or equal to β. Moving the value of β toward -1 will increase the number of lines in CLL_E (decrease $|CLL_u|$ and increase simulation speedup); however it may affect the TEP result as some effective candidate lines may be removed form the candidate line list. These tools provide more flexibility and control for planners to choose a fraction of CLL. Depending on the whole optimization process, one may prefer to get a sub-optimal answer but faster in this step, and modify it in the next steps. In this chapter, we keep β close to zero.

The proposed method is heuristic so we cannot guarantee the optimality of results when CLL_u is used instead of CLL_o. However, in most real cases this sub-optimality is acceptable as it helps to get the answer in a reasonable time.

3.4 Case study and Simulation results

All results in this section have been obtained from a personal computer with 2.0-GHZ CPU using MATLAB R2014a [74] and YALMIP R20140221 package [60] as a modeling software and GUROBI 5.6 [45] as the solver.

The ERCOT model introduced in section 2.4.2 is used for simulation. We considered transmission investment as having an overnight cost of \$1M/mile (16% of this cost is considered as annual investment cost). Value Of Loss Load (VOLL) is \$9000/MWh and wind curtailment penalty is set to \$500/MWh. It is also assumed that all scenarios have the same probability, $\beta = -0.05$ and $\eta = 50\%$. Scenarios are created to capture uncertainties in wind and load [94].

Three case studies i.e case A with one scenario, case B with 5 scenarios, and case C with 10 scenarios are analyzed to evaluate the impact of uncertainties in wind and load on both TEP and CLL_u. Based on [94], it is assumed that all network lines can be reinforced, which means $|CLL_o| = 427$, and it covers the whole geographical area of the existing network. It is only for demonstration purpose, and in real cases different criteria may be used to create CLL_o for a planning horizon. Solving TEP with this large number of binary variables is hard, and computational time increases significantly with increasing number of scenarios.

The steps of the proposed method are: after loading input data, islanding is checked, and no island is found. Relaxed OPF is solved for all three cases, and MLL is created for each case based on load and wind availability in each case. Size of CLL_E and CLL_u for three cases are given in Table 3.1. For case A, the size of CLL_u is 89.4% less than the size of CLL_o ($1 - \frac{|CLL_u|}{|CLL_o|} = 0.894$), and 78.9% for case B and 77.75% for case C. This reduction will significantly reduce the computational time for solving TEP. From the second row in Table 3.1, the number of candidate lines is still large compared to previous literature. It is still possible to reduce this list more by considering environmental constraints, but this is not in the scope of this dissertation.

After creating CLL_u for all three cases, TEP is solved for each case. The selected lines and computational time are shown in Table 3.2. For case A, we need to build 2 new lines and 4 new lines for case B and C based on the second row in Table 3.2. Based on these results, considering more scenarios may affect the selected expansion plan. So, decreasing computational time by reducing the size of CLL provides the opportunity for more accurate uncertainty modeling.

To evaluate the performance of the proposed method (that uses CLL_u as its candidate line list), the original TEP problem with all initially proposed candidate lines (CLL_o) is also solved. Both methods select the same optimal expansion plan for all these three cases (selected lines are shown in Table 2.8). For three cases A, B and C, the ratio of simulation run time for the original and the proposed method is shown in Figure 3.3. For case A, the simulation time for the original method is 10.3 times more than the proposed method, and this ratio is 134.1 for case B and 1153 for case C. These results show that the

Table 3.1: Size of CLL for Different Cases

	Case A	Case B	Case C
Size of CLL_E	382	337	332
Size of CLL_u	45	90	95
CLL size reduction	89.4%	78.9%	77.75%

Table 3.2: Transmission Expansion Planning for Reduced ERCOT System

	Case A	Case B	Case C
Selected Lines	7670 - 7668	1131 - 1064	1131 - 1064
	90002 - 5915	1067 - 1315	60042 - 60040
	-	60042 - 60040	60044 - 60040
	-	90002 - 6009	90002 - 6009
Total Time (sec)	0.78	4.8	11.2

relative performance of the proposed method increases with increasing number of scenarios (increasing the problem size). Lower values of β (close to -1) will result in a shorter candidate lines list that reduces computational time, but it may affect the selected plan. Therefore, there is a trade-off between speedup and the accuracy of results.

3.5 Summary

In this chapter, a trackable heuristic method is proposed to decrease the size of candidate lines list (CLL) in high level transmission capacity expansion planning for large scale networks. By running a relaxed OPF, lines with violations are detected and are added to monitored lines list (MLL). Line Closure Distribution Factor (LCDF) is calculated for each candidate line to evaluate

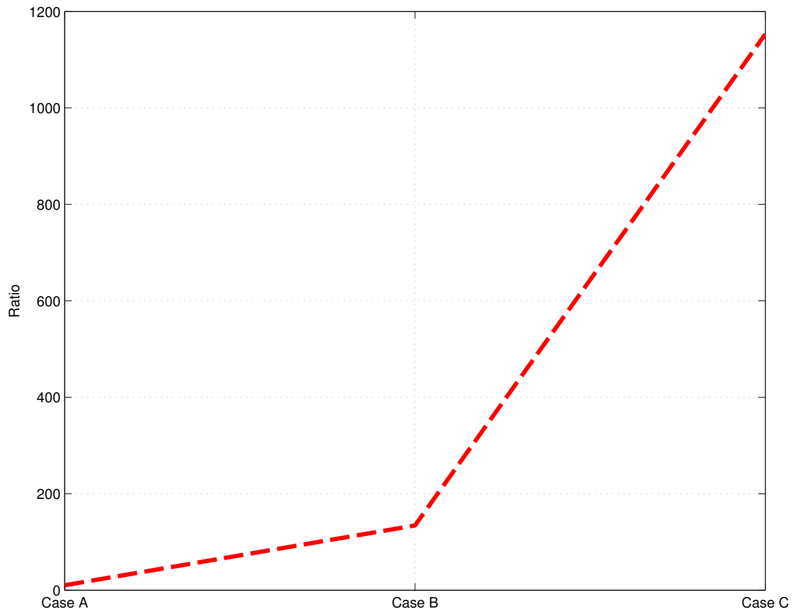


Figure 3.3: The simulation run time ratio

the impact of its closure on flow on lines in MLL. Lines that their closure will increase flow in all monitored lines are removed from CLL, because those lines will have negative impact on congestion mitigation in the base network. TCEP is solved with updated CLL. The simulation results show the effectiveness of the proposed method on reducing the computational time (more than 1100 times faster for reduced ERCOT case with 317 buses, 427 candidate lines and 10 scenarios). It allows planners to consider more accurate models for transmission planning like adding more scenarios for better representation of uncertainties as it may affect the optimal plan.

Chapter 4

Special Scenario Selection for Stochastic TEP with Contingency Analysis[1]

[1]M. Majidi-Qadikolai and R. Baldick. Stochastic transmission capacity expansion planning with special scenario selection for integrating n-1 contingency analysis. IEEE Transactions on Power Systems, 31(6):4901-4912, Nov 2016. Authors had equal contributions.

Nomenclature

Sets and Indices:

N_b: Set of buses; index i, k, n

N_g: Set of all generators; index g

N_o: Set of all existing lines; index l, m

N_n: Set of all candidate lines; index l, m

N_u^{υ}: Set of all existing lines and selected candidate lines; index l, m

L_k: Set of lines connected to bus k

G_k: Set of all generators connected to bus k

W_k: Set of wind generators connected to bus k

$\Phi_l^{\omega,\upsilon}$: Set of lines with violated post-contingency flows under outage of line l in scenario ω

Ω: Set of scenarios; index ω

$N_s^{\omega,\upsilon}$: Set of system operation states under scenario ω; index c ($c = 1$ represents the normal operation condition)

$ICL^{\omega,\upsilon}$: Set of important lines for contingency analysis in scenario ω

υ: Superscript/index for iteration number

$|\ \ |$: Size of a set

Parameters :

q_i: Per MWh load shedding penalty at bus i

γ_i: Per MWh wind curtailment penalty at bus i

Co_g: Per MWh generation cost for generator g

ζ_l: Annual cost of line l construction

d_i^{ω}: Demand at bus i in scenario ω

P_g^{max}: Maximum capacity of generator g

P_g^{min}: Minimum capacity of generator g

f_l^{max}: Maximum capacity of line l

f_l^{min}: Minimum capacity of line l

M_l: Big M is a large positive number for line l

$C^{\omega,v}$: Matrix of contingencies (operation states) that specifies the status of lines under different contingencies (1 for in service and 0 for out of service lines) for scenario ω; index$c$

$\Gamma_{m,l}^{\omega,v}$: Magnitude of violation in flow of line m when line l is on outage in scenario ω

$CII_l^{\omega,v}$: Contingency identification index for outage of line l in scenario ω

δ: Relaxation factor for $CII_l^{\omega,v}$

α: Line capacity modification factor for contingency conditions ($Emergency\,capacity\,Rating = (1+\alpha) \times Normal\,capacity\,Rating$)

Random Variables :

$\tilde{\xi}^1$: load in MW

$\tilde{\xi}^2$: Available wind output in MW

Decision Variables :

x_l: Binary decision variable for line l

$r_{i,c}$: Load curtailment at bus i under operating state c

CW_i: Aggregated wind curtailment at bus i

p_g: Output power of generator g

$f_{l,c}$: Power flow in line l under operation state c

$\theta_{i,c}$: Voltage angle at bus i under operating state c. $\Delta\theta_{l,c}$ is voltage angle difference across line l under operating state c. $\Delta\theta_{l,c}$= $\theta_{k,c}$-$\theta_{n,c}$ for line l from bus k to bus n.

4.1 Introduction

As discussed in Chapter 3, with increasing penetration of intermittent renewable resources, uncertainty in both power system operation and planning increases. Ignoring these uncertainties in transmission capacity expansion planning (TEP) can result in over or under investment, and will affect system reliability and operation costs. However, integrating uncertainties into TEP makes this large-scale non-convex optimization problem even larger and more complex. To make it a solvable optimization problem, different simplifications are applied. In this chapter, we formulate TEP for one planning horizon (static planning), which is a subproblem of dynamic planning that considers multiple planning horizons (for example planning for next three horizons 10, 20, and 30 years).

Integrating uncertainties and reliability studies into the TEP optimization problem makes it very large and almost unsolvable for large-scale power systems. Authors in [84, 87, 18] evaluated the impact of ignoring uncertainties on transmission planning by comparing the results of deterministic, heuristic, and stochastic TEP for different case studies. Their result shows that stochastic TEP may select some lines that will not be selected by either deterministic or heuristic methods. In [5], authors integrated uncertainties and risks in load, availability of generation and transmission lines into a stochastic generation and transmission capacity expansion planning, and formulated the problem as a non-linear mixed-integer optimization problem. A probabilistic method for capturing uncertainties in TEP is proposed in [16]. They developed probabilistic locational marginal pricing (LMP) index, and suggested value-based criteria i.e. decreasing congestion cost and reducing weighted deviation of mean of LMPs for selecting new transmission lines. In [121], Benders decomposition with aggregated multi-cuts is used to solve TEP under uncertainties. Authors in [99] used Least-Square Monte Carlo dynamic programming to solve stochastic TEP. They deployed sensitivity analysis to determine decision regions to execute, postpone, or reject transmission investment candidates.

Contingency analysis is also added to TEP optimization problem as an important aspect to meet reliability standard requirements. In [120], the network model is improved by adding linear approximation of reactive power, off-nominal bus voltage magnitudes and network losses. They also integrated $N-1$ contingency analysis into TEP as a sub-problem. Authors in [1] integrated Available Transmission Capacity (ATC) constraints into a multi-stage stochastic TEP problem. They used GAMS/SCENRED as a tool to reduce randomly generated scenarios (very large number of scenarios) and solved TEP with all contingencies for IEEE-24 bus system. The impact of adding ATC constraint on TEP is evaluated; however, the performance of the model for large-scale systems is not discussed. Authors in [35] modeled stochastic TEP as a bi-level optimization problem, in which in the upper-level investment for transmission expansion is minimized and, in the lower-level, social-welfare is maximized given the expansion decisions from the upper-level problem. They used the dual of the lower-level problem to convert the problem into a single level optimization problem. They modeled outage of a pre-defined list of lines as different scenarios in the optimization problem. In [17], transmission expansion and reinforcement is formulated as a stochastic optimization problem to reduce vulnerability of the system in case of deliberate attacks. They developed a set of scenarios to model different plans for destroying a set of transmission lines. Authors in [55] proposed stochastic flexible transmission planning by considering adding phase shifter or non-network options such as energy storage devices and demand response. They used Benders decomposition to solve this problem. They applied the proposed model on the IEEE-RTS case with 24 buses, and the performance of the method for large-scale networks is not evaluated. In [57] and [49], the authors provided a comprehensive review of different methods for transmission expansion planning.

As discussed in Chapter 2, it is not necessary to explicitly integrate a single outage of all lines into TEP to satisfy the $N-1$ criterion, and an algorithm was developed in Chapter 2 to decrease the list of important contingencies that should be modeled explicitly, and thereby reduce computational time for a deterministic TEP optimization problem. In Chapter 3, stochastic TEP was formulated as a mixed-integer linear optimization problem, and a heuristic method was developed to reduce the number of candidate lines (the number of binary variables) to decrease the computational time for large-scale problems. Although using algorithms developed in chapters 2 and 3 reduce computational time, adding $N-1$ contingency analysis into stochastic TEP can easily result in another unsolvable problem. In this chapter, we propose a new framework that adds reliability constraints gradually to solve stochastic

TEP optimization problem with $N-1$ contingency analysis for larger-scale systems.

The rest of this chapter is organized as follows: in section 4.2, the main concepts and the proposed optimization framework are explained. The mathematical formulation of stochastic TEP, updated VCL formulation and the three-level filtering algorithm are presented in section 4.3. In section 4.4, the proposed method is applied to different case studies. Section 4.5 has a summary of the chapter.

4.2 Proposed Optimization Process

4.2.1 Integrating Expert Knowledge with TEP

Using expert knowledge (EK) for solving large-scale TEP optimization problem is inevitable with current existing machines and software. But there are different points of view on how EK should be integrated into the transmission planning decision making process. In one approach, decisions are mainly made by experts based on their expertise instead of using an optimization based method. A second approach integrates EK into the TEP decision making process as input data for an optimization problem such as the worst case scenario for planning, list of possible contingencies, a reduced list of candidate lines and so on. A third approach converts EK to some criteria (where applicable), and tries to integrate them into the TEP optimization problem. Compared to the second approach, this method is systematic and tractable on the one hand, and more challenging from the modeling perspective on the other hand. The fourth approach tries to use EK as little as possible, and solve the problem through pure mathematical formulation. These purely mathematically driven methods are usually computationally very expensive, and are not practical for large-scale problems. Authors in [88] explained that the current practices for TEP in the United States mostly follow approaches one and two.

In this chapter, we tried to move TEP decision making process from approach two to three by developing a new framework that makes it possible to integrate EK into the TEP optimization problem in a tractable way.

4.2.2 Main Concepts Description

In this subsection, concepts that mainly affect our TEP modeling and the proposed method are explained. These concepts include long vs. near-term planning, how uncertainties are modeled, and the purpose and main tasks of the filter along with different components that are involved in the design of the filter i.e. the VCL algorithm, important scenario identification index and similar scenario elimination technique.

4.2.2.1 Long-term vs. near-term TEP

By introducing new technologies, developing smart grids, flexible transmission operation and wide area monitoring systems, system operators will have more flexibility in real-time operation, and can take several corrective actions to operate power systems reliably. Decisions regarding adding these components to the transmission network are usually made in near-term TEP, in which "corrective expansion plans" such as installing special protection schemes, phase shifters, FACTS devices, PMUs and expansion of existing substation (by increasing transformer and/or circuit breaker capacity and so on) and existing transmission lines (by reconductoring or double circuiting currently single circuit lines) are proposed to improve power system reliability. These near-term expansion plans usually can be implemented in less than 5 years.

On the other hand, in long-term TEP (which is the main focus of this chapter), decisions regarding building new EHV transmission lines, substations, or increasing the highest voltage level of the network (for example an increase from 345kV to 765kV) are made. Implementing long-term expansion plans usually takes more than 5 years. For system operation modeling in the long-term TEP, day-ahead unit commitment/dispatch is used without integrating corrective actions mainly because most of these tools have settings that are sensitive to the current network configuration (for example special protection system), and need to be revised/validated after expanding the network and changing network configuration. Moreover, these extra flexibilities are usually considered as transmission network reserve for real-time operation, in which system should be reliable for $N-1-1$.

4.2.2.2 Uncertainties and scenarios

Due to increasing environmental concerns, permitting and building transmission lines takes longer, and it raises the need for longer-term TEP that increases uncertainties [68]. Uncertainties can be categorized as micro uncertainties, which are mostly related to variations in load and wind (modeled in [94], [1], [54]) and macro uncertainties, which are mostly related to uncertainties in long-term generation expansion, environmental and market regulation changes (considered by [84], [85]) and smart grids. From the statistical modeling perspective, uncertainties are also categorized as random and nonrandom as explained in [16] in detail.

To capture all these uncertainties, usually a large number of scenarios are generated in the early stages of planning (there are different methods to generate scenarios to represent uncertainties such as Monte Carlo method (used by [1]) and using historical data with statistical modeling (used by [94])), and different clustering techniques are developed to reduce the number of scenarios [87], [94]. There are also some commercial packages such as SCENRED (by GAMS group) that can be used for this purpose (used by [1]). In this paper, we consider wind availability and load variations as uncertain parameters, and historical data with statistical modeling is used to generate scenarios to capture uncertainties in wind and load for stochastic TEP. It is assumed all scenario reduction techniques are already applied, and we have a set of scenarios (Ω) that should be integrated into TEP to capture uncertainties in the future. The type of uncertainty and how it is modeled will affect the selected expansion plan in TEP. However, we are not here concerned about the origin of scenarios because the proposed iterative framework with the designed filtering algorithm for integrating contingency analysis into stochastic TEP is applicable for different scenario generation techniques.

4.2.2.3 The Filter

As stated in section 4.2.1, using expert knowledge can be very helpful for reducing computational time for large-scale problem solving, but when uncertainty increases it will be much harder (and less tractable) to directly use expert knowledge in the decision making process. For integrating contingency analysis into the TEP problem, we developed a filtering mechanism to select a subset of important lines for contingency analysis to be integrated into stochastic TEP in each iteration instead of asking experts to manually choose some lines for these analysis. The filter uses an updated version of the VCL algorithm (proposed firstly in Chapter 2) and two new indices developed (explained in the following subsections) to select a subset of scenarios and lines for contingency analysis. The advantage of this filter is that it provides a systematic and tractable way for integrating contingency analysis into TEP optimization problem gradually. More detail about the filter is given in sections 4.2.3 (step 7) and 4.3.3.

4.2.2.4 Updated VCL algorithm

The developed VCL algorithm in Chapter 2 finds all important lines for contingency analysis (ICL^{ω}s), and integrates them into the TEP at once. But for large-scale stochastic TEP problems, the size of contingency list ($|CL| = \sum_{\Omega} |ICL^{\omega}|$) will increase rapidly and makes the TEP optimization problem unsolvable or extremely computationally expensive. In this chapter we add two new features to the VCL algorithm that will let us select a subset of important lines for contingency analysis. The first one is the relaxation factor δ that selects a subset of lines with high contingency identification index (see section 4.3.2 for more detail), and the second one is the ability to select a fixed fraction of lines that adds more flexibility to managing the size of contingency list.

4.2.2.5 Important Scenario Identification Index (ISII)

Different scenarios affect power system operation differently under normal operation condition (for example, more power plants will be committed/dispatched when demand is high compared to low load condition at midnight). Under contingency operation states, the VCL algorithm will select different lines under different scenarios, and the size of ICL^{ω} may significantly change from one scenario to another.

We define a set of scenarios "normal" for *contingency analysis* if its contingency statistics vector (referred as CS), which contains the number of important lines for contingency analysis in each scenario ($|ICL^{\omega}|$s), has a normal distribution (there are different tests to check normality of a distribution such as Kolmogorov-Smirnov, Lilliefors, and Jarque-Bera [110]). Based on this definition, we set $ISII = 0$ for a scenario set with CS having a normal distribution with a small standard deviation. It means $|ICL^{\omega}|$s are mostly close to the mean of the set with a few far from it that show a normal behavior of the scenario set from contingency analysis perspective; therefore there is no special scenario in this set to be evaluated separately. Otherwise, $ISII$ is set to one ($ISII = 1$) that shows there are some scenarios that have significantly different behavior compared to the average in the set from the contingency analysis perspective; so we would like to separate them from the rest and analyze them separately.

To find important scenarios for the case with $ISII = 1$, a normal distribution is fitted to CS vector, and scenarios with $|ICL^{\omega}|$ larger than *mean* plus *one standard deviation* of the fitted normal distribution are tagged as important scenarios, and are stored in Important Scenarios List (ISL). If the ISL is not empty, lines for contingency analysis will be selected from lines in ICLs of the scenarios in ISL. It will result in a short list of more effective lines for contingency analysis in the first iteration.

4.2.2.6 Similar Scenario Elimination (SSE) technique

By using $ISII$, we separated specific scenarios from the rest. What can we say about scenarios such that their ICL^{ω}s have relatively the same size? Can we assume they are all the same? We cannot answer this question only based on the size of their important contingency lists, because there might be totally different lines in each ICL^{ω}. For example in a high-wind/low-load scenario different lines might be selected by the VCL algorithm compared to a low-wind/high-load scenario; therefore, we need to look at lines in ICL^{ω} of each scenario ω to be able to compare them. When two scenarios have relatively similar lines in their ICLs, we can eliminate one of them from contingency analysis as their impact on contingency analysis will not be significantly different. The Similar Scenario Elimination (SSE) technique is developed based on this concept. ICL^{ω} is a list that contains important lines for contingency analysis for scenario ω. In $ISII$, a vector of $|ICL^{\omega}|$s (called CS) is used to make decision about scenarios, and in SSE we look inside these lists to make a decision. SSE checks the similarity of lines in ICLs of a scenario set/subset to find scenarios with more than a specific percent of similarity in their lists. Then, among similar scenarios, one with a greater number of important lines will be selected to create contingency list (CL) vector based on its ICL^{ω}s, and ICLs of other similar scenarios will be eliminated from contingency analysis in that iteration. SSE can be applied to scenario sets with both $ISII = 0$ and $ISII = 1$ to decrease the number of lines for contingency analysis.

It should be emphasized that we do not remove any scenario from stochastic TEP, and the size of operation states set for scenario ω in iteration υ, which is represented by $N_s^{\omega,\upsilon}$, is always greater than or equal to 1 ($|N_s^{\omega,\upsilon}| \geq 1\ \forall\omega,\ \forall\upsilon$). In other words, in each iteration that TEP optimization is solved (in step 9 of the proposed framework) all scenarios are included in the optimization problem at least for their normal operation state. We create CL from ICLs of a subset of scenarios by using the designed three-level filter to reduce computational time in early iterations. However, the iterative framework is terminated if and only if all important lines for contingency analysis are integrated into the TEP optimization problem; therefore the contingency list (CL) will contain all ICLs of all scenarios in the last iteration.

4.2.3 The Proposed Framework

In this chapter, a framework is designed to iteratively solve stochastic TEP with $N-1$ contingency analysis. The proposed three-level filter is used to select a subset of important reliability constraints for the optimization problem in each iteration to increase the problem size gradually and thereby reduce overall computational time compared to considering all constraints explicitly from the start. The proposed framework is summarized in the following 10 steps:

Step 1: Load input data, set $\upsilon = 1$.

Step 2: Check system islanding

In this step, an algorithm checks for island buses in a network in which all candidate lines are tentatively built. If any island buses are found in this step, they will be deleted from data permanently as they will never be connected to the network.

Step 3: Solve a relaxed version of TEP

In this step, all constraints related to contingency analysis are ignored, and a relaxed version of the original integrated TEP is solved. This optimization problem is much easier to solve and provides a lower bound for the original problem. The existing network (N_o) is updated by adding the selected candidate lines to create updated network N_u^v. The new system is referred to as S_u^v.

Step 4: Temporarily remove island buses

If there is any island bus in the updated system (S_u^v), it will be removed temporarily from data as it will not have any impact on ICL and filtering. The new reduced system is called S_r^v.

Step 5: Create ICL for all scenarios

For S_r^v, $ICL^{\omega,v}$ will be created for all scenarios with relaxation factor value $\delta = 1$. Mathematical formulation and full definition of relation factor are given in section 4.3.2.

Step 6: Create Contingency Statistics (CS_o^v) and Contingency Lines (CL_o^v) vectors

For this step, the contingency statistics vector that contains the size of $ICL^{\omega,v}$s for each scenario ($|ICL^{\omega,v}|$s) is created and labeled CS_o^v, and all important lines for contingency analysis are added to the contingency list (CL_o^v) vector.

Step 7: Three-Level Filtering for contingency analysis

A three-level filter is designed to further decrease the total number of lines for contingency analysis in TEP based on network and scenario set characteristics. In each iteration only one level of the filter will be selected to modify CS_o^v and CL_o^v to form vectors CS^v and CL^v to be integrated into TEp optimization problem in step 9.

- *High-Level Filter*: The algorithm gets into this level if $ISII = 1$ and $\upsilon = 1$. After running $ISII$ and creating ISL, if $|ISL| = 1$, the updated VCL algorithm is run with a δ value close to 1. If $|ISL| > 1$, SSE algorithm is run to eliminate similar scenarios in ISL if there are any. Therefore, a relatively small subset of lines in CL_o^υ are selected at the end of this level of filtering, and CS^υ and CL^υ are created.

- *Medium-Level Filter*: The algorithm gets into this level in an iteration if it did not get into the previous level and the *mean* of CS_o^υ is large enough. In this level, SSE is used to find and eliminate similar scenarios. If the number of remaining lines for contingency is still large, the updated VCL algorithm with $\delta \gg 1$ is run to select a fraction of lines to reduce the size of contingency lines list. At the end of this level CS^υ and CL^υ are created.

- *Low-Level Filter*: If the algorithm did not get into the first or the second levels in an iteration, it will get into this level. In this step only the updated VCL algorithm with $\delta \geq 1$ will be run to reduce the size of contingency lines list and create CS^υ and CL^υ.

The filter is designed in a way to ensure that sum of CS^υ elements in iteration υ is greater than or equal to iteration $\upsilon - 1$. Otherwise $CS^\upsilon = CS_o^\upsilon$ and $CL^\upsilon = CL_o^\upsilon$, which means all important lines will be selected for contingency analysis (all scenarios are included in contingency analysis).

Contingency matrices $C^{\omega,\upsilon}$ are created based on CS^{υ} and CL^{υ} in each iteration. See section 4.3.3 for more detail.

Step 8: Check Stopping Criteria

The iterative process will stop when $CS^{\upsilon} = CS_o^{\upsilon}$ in an iteration. In this case, the variable $Flag$ is set to 1.

Step 9: Solve TEP optimization problem

In this step, S_u^{υ} is used as the base case for TEP. After solving TEP optimization problem, set $\upsilon = \upsilon + 1$, and update N_u^{υ} and S_u^{υ} based on the selected plan in this step. If $Flag = 1$ go to step 10, otherwise return to step 4.

Step 10: We have the optimal/near-optimal expansion plan that satisfies $N-1$ criterion.

We can confirm that the selected plan satisfies the $N-1$ criterion by running a DC-SCOPF with all contingencies for the selected expansion plan to make sure there is no violation in constraints, and if there is any, the algorithm will return to step 4 to update CS and CL. As selected lines in each iteration are considered as the built lines for the next iteration (by updating N_u^{υ} and S_u^{υ} at each iteration), in general we cannot guarantee global optimality of the final result. To quantify the quality of the final result, we need to calculate the optimality gap by finding a lower bound answer (discussed in section 4.2.4). The proposed framework is summarized in a flowchart in Figure 4.1.

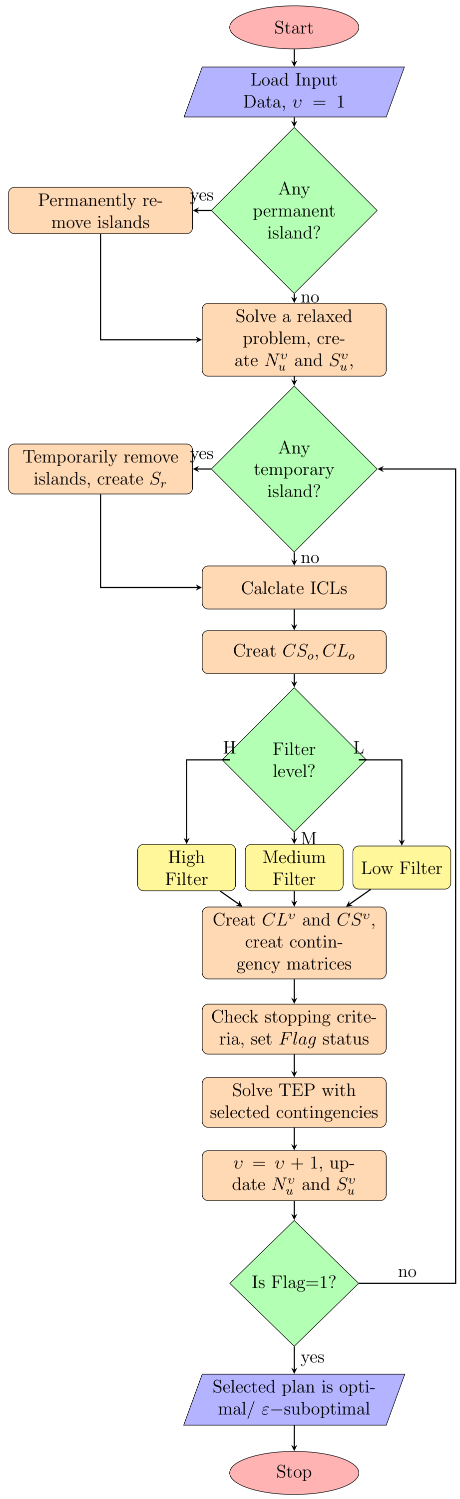


Figure 4.1: Flowchart of the proposed iterative framework

4.2.4 Sub-Optimality Bound

Branch and Bound (BB) is one of the most common methods for solving combinatorial optimization problems. It was proposed by A. H. Land and A. G. Doig [56] and improved by several references since 1960 [20]. To quantify the quality of the obtained result by the proposed method, we use BB to find a lower bound (LB) for TEP by exploring some levels of branches for improving the lower bound. Details of BB algorithm is not in the scope of this paper. The result of step 10 is considered as the upper bound (UB) for calculating the error from the following equation:

$$\varepsilon = \frac{UB - LB}{LB} \times 100 \tag{4.1}$$

The value of ε in (4.1) depends on the number of applied branch and bound steps for calculating the lower bound answer, and it shows the maximum possible error between the answer in step 10 and the globally optimal answer (also called $\varepsilon-$suboptimal). As the proposed iterative method always selects a subset of the most important reliability constrains for solving TEP in each iteration, we expect the optimal answer to be close to the upper bound.

4.3 Mathematical Formulation

4.3.1 Two-Stage Stochastic TEP Formulation with Dynamic Contingency Matrix

In our formulation, we assumed wind and load as uncertain parameters, but as explained in section 4.2.2 the proposed method is independent of the nature of uncertainty and the origin of scenarios (as long as uncertainties can be represented by scenarios with known probabilities). When probabilities are unknown, their values can be estimated using Bayesian approaches, where non-informative prior probabilities are established in correspondence with the objective variable of the problem (see [39, 40] for more detail) .Compared to Chapter 2, in this formulation, the deterministic operation cost is replaced with the expected operation cost that captures the impact of future uncertainties on system operation. Compared to Chapter 3, contingency analysis is added to stochastic TEP to guarantee the selected plan meets $N-1$ criterion. The two-stage stochastic TEP is formulated as follows:

$$Z^{*}=\min \sum_{N_n} \zeta_l x_l + \mathbb{E}[Q(\boldsymbol{x}, \tilde{\xi})] \tag{4.2}$$

$$\text{st.} \quad x_l \in \{0,1\} \quad \forall x_l \in N_n, \tag{4.3}$$

where $\tilde{\xi}$ is a random variable vector that represents load and wind uncertainties ($\tilde{\xi} = (\tilde{\xi^1}, \tilde{\xi^2})$). $\mathbb{E}[Q(\boldsymbol{x}, \tilde{\xi})]$ represents the expected value of operation costs that contains load shedding and wind curtailment penalty and generation costs. This expected value is approximated with a weighted sum of a limited number of scenarios as follows [31]:

$$\mathbb{E}[Q(\boldsymbol{x}, \tilde{\xi})] \approx \sum_{\Omega} P^{\omega} Q(\boldsymbol{x}, \xi^{\omega}) \tag{4.4}$$

where $Q(\boldsymbol{x}, \xi)$ is the optimal value of power system operation for a given load and wind represented by ξ.

$$Q(\boldsymbol{x}, \xi) = \min \sum_{N_s} (\sum_{N_b} q_i r_{i,c}) + \sum_{N_b} \gamma_i CW_i + \sum_{Ng} Co_g p_g \tag{4.5}$$

$$\text{st.} \quad -\sum_{L_k} f_{l,c} + \sum_{G_k} p_g + r_{k,c} = d_k \tag{4.6}$$

$$-M_l(1 - C_{l,c}x_l) \leq f_{l,c} - B_{l,l}\Delta\theta_{l,c} \tag{4.7}$$

$$M_l(1 - C_{l,c}x_l) \geq f_{l,c} - B_{l,l}\Delta\theta_{l,c} \tag{4.8}$$

$$CW_k \geq \sum_{W_k} (P_g^{max} - p_g) \tag{4.9}$$

$$(C_{l,c}x_l) f_l^{min} \leq f_{l,c} \leq f_l^{max}(C_{l,c}x_l) \tag{4.10}$$

$$P_g^{min} \leq p_g \leq P_g^{max} \tag{4.11}$$

$$0 \leq r_{k,c} \leq d_k \tag{4.12}$$

$$-\frac{\pi}{2} \leq \theta_{k,c} \leq \frac{\pi}{2} \tag{4.13}$$

$$CW_k \geq 0 \tag{4.14}$$

$$x_l = 1, \quad \forall l \in N_o \tag{4.15}$$

$$x_l \in \{0, 1\}, \quad \forall l \in N_l \tag{4.16}$$

This optimization problem is solved in step 9 for every iteration. In equation (4.5), load shedding is penalized over all operating states (N_s) to satisfy the $N-1$ criterion (no load shedding is accepted during both normal and under single contingency states). Equation (4.6) enforces power balance at each bus. Equations (4.7) and (4.8) show DC representation of flow in transmission lines with big M technique. Equation (4.9) measures wind curtailment at each bus. Equation (4.10) shows flow in all lines should always be between their maximum and minimum capacity limits. These limits will be modified based on the given value for α for emergency conditions (contingency in the network). Equations (4.11)-(4.13) enforce power plants' dispatch, load shedding and voltage angles to be between their minimum and maximum limits. Equation (4.14) enforces non-negativity of wind curtailment. Equation (4.15) sets decision variables for existing lines to 1. Equation (4.16) enforces that $\boldsymbol{x}$ is a binary decision variable for transmission lines ($x = 1$ when a line is built and $x = 0$ when a line is not built).

4.3.2 Updated Variable Contingency List (VCL) Algorithm

Modified Line Outage Distribution Factor (LODF) is used to calculate post-contingency flow in transmission lines when one line is on outage. The following equations are used to create important contingency lists for different scenarios.

$$\Gamma_{m,l}^{\omega,v} = \frac{f_{m,l}^{\omega,v} - f_m^{max}}{f_m^{max}} \qquad , \forall m, l \in N_u^v, \forall \omega \in \Omega \tag{4.17}$$

$$\Phi_l^{\omega,v} = \{ m \in N_u^v \, | \Gamma_{m,l}^{\omega,v} \geq \alpha \} \quad , \forall l \in N_u^v, \forall \omega \in \Omega \tag{4.18}$$

$$CII_l^{\omega,v} = \begin{cases} \frac{\sum_{\Phi_l^{\omega,v}} \Gamma_{m,l}^{\omega,v}}{|\Phi_l^{\omega,v}|}, & if \quad |\Phi_l^{\omega,v}| \neq 0 \\ 0, & if \quad |\Phi_l^{\omega,v}| = 0 \end{cases} \tag{4.19}$$

$$ICL^{\omega,v} = \{ l \in N_u^v \, | CII_l^{\omega,v} \geq \alpha\delta \} \quad , \forall \omega \in \Omega \tag{4.20}$$

$$\delta \geq 1, \tag{4.21}$$

where (4.17) calculates over/under loading on line m when line l is out. In this equation, $f_{m,l}^{\omega,v}$ represents the magnitude of post-contingency flow in line m when line l is on outage. Equations (4.18)-(4.19) are used to calculate Contingency Identification Index (CII) for each scenario with α as the line capacity modification factor during contingencies (see [69] for more details). Equation (4.20) creates important contingency list (ICL) based on CII. To be able to select a fraction of lines in ICLs, a relaxation factor δ is included in (4.20). This new capability is useful for managing the size of CL in different iterations.

4.3.3 Three-Level Filtering Algorithm

Algorithm 1 explains the proposed three-level filter in step 7 in section 4.2.3 in more detail. To develop this algorithm, concepts explained in section 4.2.2 are used.

As shown in Algorithm 1, after checking the normality of CS_o^v distribution in iteration v (H), mean (μ) and standard deviation (σ) of the fitted normal distribution is calculated. Then the status of $ISII$ will be set (based on H, μ, σ). In the next step, conditions for selecting a filter level is checked. For the high-level filter, first the ISL is created, then based on the number of important scenarios ($|ISL|$), the filter goes through SSE or the updated VCL algorithms. For the medium-level filter, SSE and the updated VCL algorithms (if applicable) are run to reduce contingency lines list. The low-level filter applies the updated VCL algorithm to create CL^v and CS^v. In each it-

Algorithm 1 Three-Level Filtering

Require: CS_o^v, CL_o^v *and* v

$H \leftarrow 0$ $(if\ CS_o^v\ is\ not\ Normal)$

$\mu \leftarrow mean\ of\ CS_o^v$

$\sigma \leftarrow standard\ deviation\ of\ CS_o^v$

$ISII \leftarrow 1$ $(if\ H = 0\ OR\ \sigma \geq \mu/2)$

if $v = 1$ AND $ISII = 1$ ***(High-Level)*** **then**

  $ISL \leftarrow Scenarios\ with\ |ICL| \geq (\mu + \sigma)$

  if $|ISL| > 1$ **then**

    $CS^v, CL^v \leftarrow Run\ SSE\ for\ ISL$

  else

    $CS^v, CL^v \leftarrow Run\ updated\ VCL\ for\ ISL$

  end if

else if $v \leq 3$ AND $\mu\ is\ large$ ***(Medium-Level)*** **then**

  $CS^v, CL^v \leftarrow Run\ SSE$

  if $sum(CS^v)\ is\ still\ large$ **then**

    $CS^v, CL^v \leftarrow Run\ updated\ VCL\ with\ \delta \gg 1$

  end if

else ***(Low-Level)***

  $CS^v, CL^v \leftarrow Run\ updated\ VCL\ with\ \delta \geq 1$

end if

Ensure: $Sum(CS^v) \geq Sum(CS^{v-1})$ OR $CS^v = CS_o^v$

eration to guarantee the algorithm's eventual termination, it is always checked that the number of selected lines increases or that all lines will be selected. It is critical to design the filter in a way that effectively creates CL^v and CS^v based on the size of the network and the number of scenarios. In section 4.4.1, the detail of applying all filtering steps for a numerical case study is given.

4.4 Case study and Simulation results

In this section, we run numerical analysis for five case studies on a 13-bus system (three of the cases) and a reduced ERCOT system (two of the cases). All simulations are done with a personal computer with 2.4-GHz CPU and 16 GB of RAM. The proposed method is implemented in MATLAB R2014a [74] by using YALMIP R20140221 package [60] as a modeling software and GUROBI 5.6 [45] as a solver. To calculate "Total Time", MATLAB built-in function `tic toc` is used to evaluate elapsed wall clock time, and "Solver Time" is calculated by YALMIP. The case studies are solved for three methods i.e. the proposed method in this paper, the VCL algorithm [69] and the integrated model (in which $N-1$ contingency analysis is modeled for all lines) and their performance are evaluated.

4.4.1 13-Bus System

This system has 13 buses, 33 existing lines, 16 power plants, 9 load centers, and 36 candidate lines. It is developed for educational purposes and represents a simplified version of the ERCOT network (see Figure 4.2). Bus

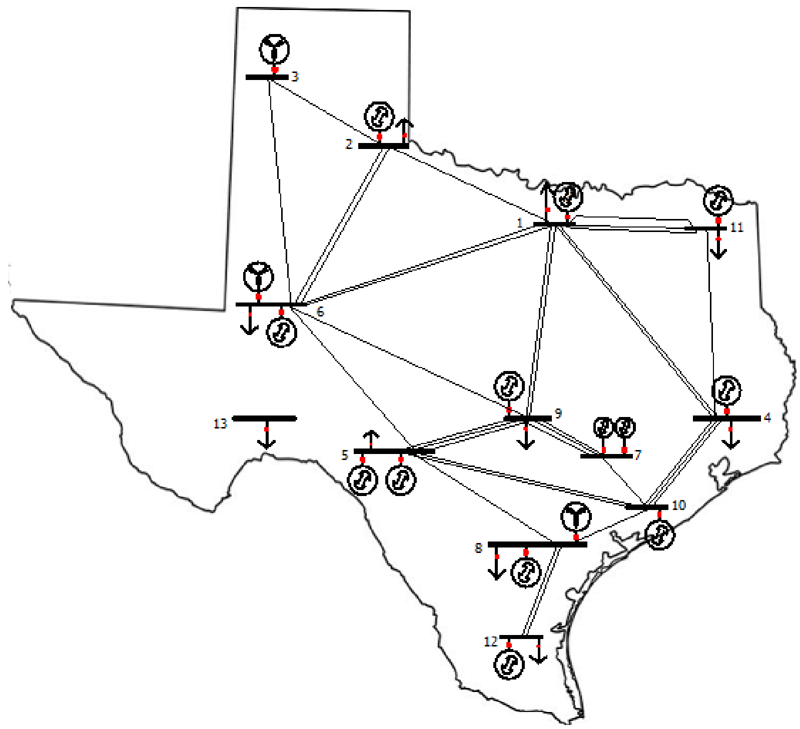


Figure 4.2: 13 bus network with existing lines, generators and loads.

No. 13 represents a new demand center that submitted a request to connect to the network. High wind expansion in west (buses 3 and 6) also introduces needs for network reinforcements in west and central parts of Texas. This system is run for three cases i.e. A, B and C with 20, 50 and 100 scenarios, respectively, to evaluate the capabilities of the proposed method. It is assumed that all scenarios have the same probability. Transmission investment cost is \$1M/mile that is converted to an annual base (16% of this cost is considered as annual investment cost). Value Of Loss Load (VOLL) is set to \$9000/MWh and wind curtailment penalty is \$500/MWh.

To show the performance of the proposed method when it deals with different cases, the results of these three cases are compared for all steps of the proposed method. Step 1: Input data is loaded, $\upsilon = 1$. Step 2: check islanding: No island. Step 3: Solve relaxed problem: for case A, 7 new lines are added, and for cases B and C, 9 new lines are added, which will significantly affect the computational time in step 9. S_u^1 and N_u^1 are created for each case. Step 4: Temporary island removing: No island, so S_r^1 is the same as S_u^1. Step 5: Create ICL for all scenarios: by setting $\alpha = 0.05$ and $\delta = 1$, $ICL^{\omega,1}$s are created for all scenarios. Step 6: Create CS_o^υ and CL_o^υ: CL_o^1 is created from ICLs. $|CL_o^1|$ for cases A, B and C are 219, 538, and 1041, respectively. These numbers show how fast the problem size will increase for a large number of scenarios even after applying the VCL algorithm. For each case, the total number of lines in CL_o^1 is equal to the sum of elements of CS_o^1.

Step 7 applies the three-level filtering. It is evaluated step-by-step and in more detail to make its impact clear. We used the MATLAB built-in function `jbtest`, which is developed based on Jarque-Bera test, to check whether a data set has a normal distribution. CS_o^1 for case A has a normal distribution, and for cases B and C, CS_o^1 does not have a normal distribution. The value of $ISII$ is set for three cases as $ISII_A = 0$, $ISII_B = 1$, $ISII_C = 1$. The histogram and fitted normal distribution on CS_o^1 vector are shown in Figure 4.3 for each case in the first iteration. Gray bars show the frequency of $|ICL^{\omega,1}|$s, and the red bell shape curve represents the normal distribution fitted to CS_o^1. Green dashed line indicates *mean* (μ) of the fitted normal distribution and blue solid line represents *mean+std* ($\mu + \sigma$). If $ISII$ is equal to 1, then scenarios on the right side of the blue line will be tagged as important scenarios. Important Scenarios List (ISL) for each case is given in the second row of Table 4.1. Case A has no important scenario, as its CS_o^1 distribution is normal and its standard deviation is less than half of its mean (see Figure 4.3(a)). For case B and C (see Figure 4.3(b),(c)) $ISII = 1$, and there are 5 and 10 important scenarios for these two cases. For the first iteration, the filter selects 108 lines for contingency analysis for case A ($\sum_{\Omega} |N_{sA}^{\omega,1}| = 128$ operation states), 70 lines for case B ($\sum_{\Omega} |N_{sB}^{\omega,1}| = 120$) and 153 lines for case C ($\sum_{\Omega} |N_{sC}^{\omega,1}| = 253$). To evaluate the impact of the proposed filter on the problem size, the total number of operation states ($\sum_{\Omega} |N_s^{\omega,1}|$) that should be considered by different methods is given in Table 4.2. With only using the VCL algorithm from [69] (without using the filter introduced in this paper), TEP with 239, 588 and 1141 operation states should be solved for these three cases, which are much harder to solve and need significantly more computational time. Based on the last row of the table, for the integrated model that considers all lines for contingency analysis, the problem size will be so large as to easily make medium and large-scale problems unsolvable. Figure 4.4 shows the impact of the filter on reducing the size of CL^v in different iterations. Blue, red, and green colors represent case A, B and C respectively. Dashed lines show the original number of important lines ($|CL_o^v|$), and solid lines show selected lines by the filter ($|CL^v|$) in each iteration. The difference between dashed and solid lines in each iteration shows the impact of the filter on reducing $|CL|$

Table 4.1: ISLs and Selected Filter Levels for 13-bus System

	Case A	Case B	Case C
ISL	−−	17, 32, 41, 42, 50	17, 32, 41, 42, 50, 57, 78, 94, 96, 97
Filter Levels	Medium Low – – –	High Medium Medium Low Low	High Medium Medium Low Low

for each case. Reducing $|CL_o^v|$ from one iteration to another is because of the developed framework that iteratively solves TEP and updates N_u and S_u. As cases B and C have important scenarios, lines in CL^1 are selected among ICLs of scenarios in their ISLs compared to case A, in which CL^1 is created from ICLs of all scenarios in the first iteration. CL_o^v and CL^v are getting closer and closer to each other from iteration v to iteration $v+1$, and this guarantees the algorithm's termination. The third row of Table 4.1 shows how the algorithm moves between the filter's levels during iterations for different cases. During two iterations for case A, the algorithm selects medium and low levels respectively. For cases B and C, it selects High, Medium, Medium, Low and Low levels. The number of iterations and filtering levels are selected based on problem characteristics that demonstrates the dynamic behavior of the filter.

In step 8, stopping criteria is checked. As shown in Figure 4.4, at iterations 2, 5 and 5, respectively, $CS^v = CS_o^v$ and $Flag$ is set to 1 for cases A, B and C, respectively. In step 9, TEP optimization problem with selected

Table 4.2: Total number of operating states in the first iteration for different case studies

	Case A	Case B	Case C
Proposed method	128	120	253
VCL algorithm	239	588	1141
Integrated model	1400	3500	7000

contingencies is solved, $v = v+1$, and N_u^v and S_u^v are updated in each iteration. The algorithm gets to step 10 after the stopping criteria is met. Selected lines and total costs are given in the second and third rows in Table 4.3. Case A adds 11 new lines, and case B and C add 16 lines to the network. The simulation time for the proposed method is given in the fourth row in Table 4.3. The difference between total time and solver time in the fourth row of this table represents the time that the filter and the modeling language need in the process of solving the TEP optimization problem. The design of the filter will affect both solver and total time. To make sure that these results satisfy $N-1$ criterion, DC-SCOPF is run for all contingencies, and no violation occurs.

To calculate ε to quantify the quality of results in step 10, a lower bound is calculated by applying a few steps in a BB algorithm. It is possible to apply more steps to get a better lower bound answer. The error bounds for three cases are given in the last row in Table 4.3. It shows 1.3%, 2.25% and 2.9% as the upper bound error for cases A, B and C respectively. However, these are not the actual error between optimal answer and our results.

To compare the actual error with ε and show the impact of the proposed

Table 4.3: Transmission Expansion Planning for 13-bus System

	Case A	Case B	Case C
Selected Lines	2-1, 2-1, 3-2, 3-2, 3-6, 6-9, 7-10, 8-12, 13-5, 13-6, 13-6	2-1, 2-1, 1-6, 1-6, 3-2, 3-2, 3-6, 4-10, 4-11, 7-10, 8-10, 8-10, 8-12, 13-5, 13-5, 13-6	2-1, 2-1, 1-6, 1-6, 3-2, 3-2, 3-6, 4-10, 4-11, 7-10, 8-10, 8-10, 8-12, 13-5, 13-5, 13-6
Total Cost ($ M)	4889	4986.4	4945.9
Solver (sec) Total (sec)	27.2 33.15	177.8 243.6	1252 1443
ε Actual Err	1.3% 0%	2.25% 0%	2.9% 0%

Table 4.4: Total simulation time for different methods [minutes]

	Case A	Case B	Case C
Proposed method	0.55	4.06	24.05
VCL algorithm	19.9	1714	15659
Integrated model	8768 (6+ days)	NA	NA

method on reducing computational time, we run these three cases with the proposed algorithm in chapter 2 and the integrated model. The actual error is shown in the last row in Table 4.3. It shows that for all three cases the actual error is zero, which means the proposed method found the optimal plan for these cases. It should be mentioned that as in real cases we do not know the optimal answer, the quality of results is quantified by the calculated ε. In other words, our answer is $\varepsilon-$suboptimal. The performance of these three different methods is compared in Table 4.4. Each row in this table shows the simulation time each method needs to solve these case studies. The ratio of the third row to the second row in this table shows how much the proposed method in this paper performs better compared to [69] for stochastic TEP. This ratio is more than 35, 420 and 650 for cases A, B and C respectively and shows the relative performance of the proposed method increases with increasing the problem size. Compared to the integrated model, the proposed method found the answer more than 15657 times faster for case A, and we could not get any answer even after 12 days for cases B and C. This great performance is achieved because of huge problem size reduction using the designed filter through the developed iterative framework. For example for case C, in the first iteration, the proposed method decreases the number of structural constraints by 85% compared to [69] and by 98% compared to the integrated model.

4.4.2 Reduced ERCOT System

A reduced model of the ERCOT system is provided in [94]. This model is developed for evaluating the impact of high penetration of wind power on west Texas network. Therefore, the west part of ERCOT network in modeled in detail, and the rest of ERCOT is simplified into three zones. This network contains 317 buses, 427 branches, 489 conventional power plants, 36 wind farms and 182 load centers. The number of candidate lines is equal to the number of existing lines. In this TEP optimization problem, there are 427 binary variables, which makes it a challenging problem to solve. All costs are set the same as the 13-bus system. We consider two cases for the ERCOT system i.e. case A with 5 scenarios and case B with 10 scenarios. Scenarios are generated using historical load and wind data [94]. For case A, scenario 5 is selected as the important scenario, and scenarios 5 and 6 are in ISL for case B. The original number of important lines ($|CL_o^1|$) is 23 and 52 respectively, and the three-level filter selects 10 and 19 lines for the first iteration in case A and B. Both cases take two iterations to converge, and go through High and Low level filters.

The selected plan, total costs, solver and total time with ε and actual error are shown in Table 4.5. The number of selected lines is 9 and 11 for case A and B respectively. ε is around 1% for the reduced ERCOT system (the answer is 1%$-$suboptimal). Both cases are also solved with the developed algorithm in chapter 2 and the integrated model to compare the results with the proposed method in this chapter. As shown in the last row in Table 4.5, actual errors for both cases are zero, which means the proposed method found the optimal expansion plan for this system. The solver and total time are shown in the fourth row. For case A, the proposed method in [69] needs 181 minutes to solve the problem (compared to 10.5 minutes required with the proposed method in this paper), and for case B, [69] needs 12 days and 6 hours and 56 minutes to solve the problem (compared to 19.3 minutes required by the proposed method in this chapter). We did not get the answer from the integrated model after 34 days (still more than 75% optimality gap). Although the number of scenarios are not large for this system (compared to the 13-Bus

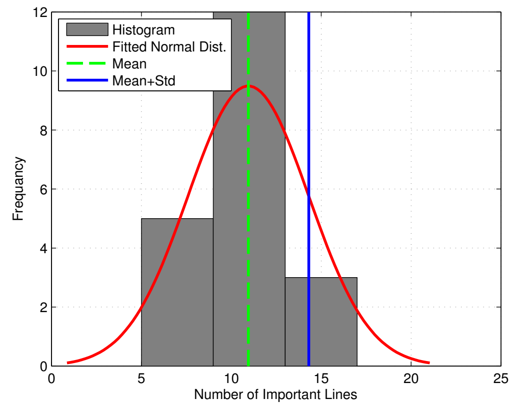
Table 4.5: Transmission Expansion Planning for Reduced ERCOT System

	Case A	Case B
Selected Lines	1311-1064, 1310-1309, 1312-1310, 1313-1312, 1315-1313, 1067-1315, 1065-1064, 1066-1065, 5905-5902	1311-1064, 1310-1309, 1312-1310, 1313-1312, 1315-1313, 1067-1315, 1065-1064, 1066-1065, 5905-5902, 60042-6216, 60042-60040
Total Cost ($ M)	16298	16857
Solver(sec)	358	871
Total (sec)	630	1160
ε	0.96%	1.01%
Actual Err	0%	0%

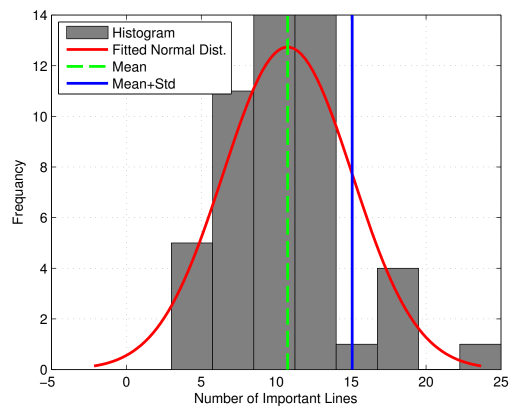
system), the proposed method gets the answer more than 910 times faster than [69], and the relative performance improves more when the number of scenarios increases. Moreover, the relative performance improvement of the proposed method grows faster for larger networks, as the ratio for the ERCOT case B with 10 scenarios is more than 27 times larger than the ratio for the 13-bus case A with 20 scenarios.

4.5 Summary

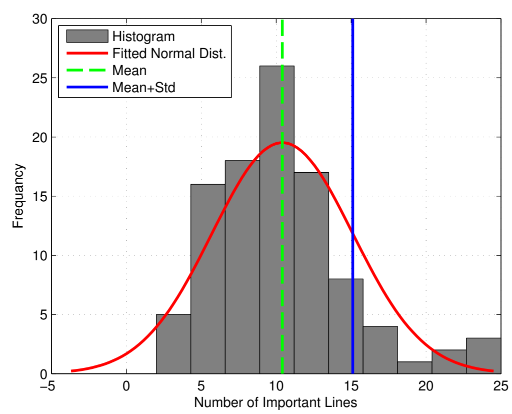
In this chapter, $N-1$ contingency analysis is integrated into stochastic transmission capacity expansion planning through an iterative framework. By developing important scenario identification index, more beneficial scenarios for contingency analysis are distinguished, and important scenarios list (ISL) is created. In the next step, the proposed three-level filtering algorithm provides a systematic, automated and tractable way to select a subset of important lines for contingency analysis. It uses ISL, developed similar scenario elimination (SSE) technique and updated variable contingency lists (VCL) algorithm to reduce the number of reliability constraints in TEP in each iteration. As an example, for the ERCOT system with 10 scenarios, the number of structural constraints decreased by 99.78% in the first iteration. The proposed method allows solution of large-scale stochastic TEP optimization problems faster by integrating contingency analysis into TEP gradually through an iterative process that decreases computational time significantly. The quality of results is quantified by calculating maximum error bound (ε optimality gap). The numerical results show that the effectiveness of the proposed method will increase by increasing the number of scenarios and size of the network. ISOs can use this algorithm as an automated tool for operation, reliability, and planning purposes.



(a) Case A



(b) Case B



(c) Case C

Figure 4.3: ISII index results for three different cases

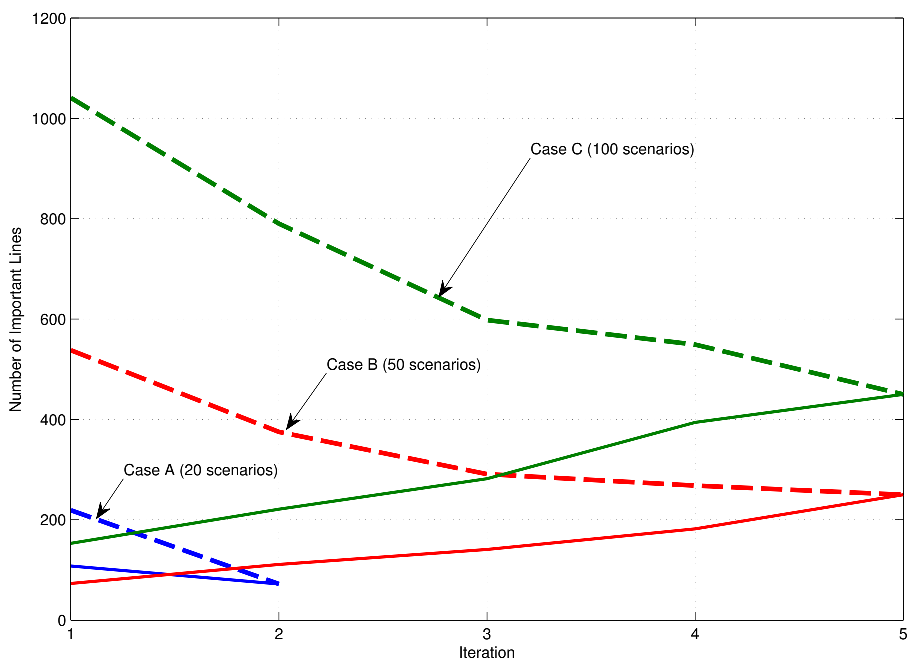


Figure 4.4: Impact of the iterative framework and the three-level filtering on $|CL_o^v|$ and $|CL^v|$ in different iterations

Chapter 5

Decomposition Framework: Concepts and Formulation[1]

[1]Mohammad Majidi-Qadikolai and Ross Baldick. A generalized decomposition framework for large-scale transmission expansion planning. IEEE Transactions on Power Systems, PP(99):1-1, 2017. Authors had equal contributions.

Nomenclature

Sets and Indices:

N_b: Set of buses; index k, n

N_g: Set of all generators; index g

N_{wg}: Set of all wind generators; index g

N_l: Set of all lines (existing and candidate); index l, m

N_o: Set of all existing lines; index l, m

N_n: Set of all candidate lines; index l, m

L_k: Set of lines connected to bus k

G_k: Set of all generators connected to bus k

Φ_l^ω: Set of lines with violated post-contingency flows under outage of line l in scenario ω

N_s^ω: Set of system operation states under scenario ω; index c ($c = 1$ represents the normal operation condition)

ICL^ω: Set of important lines for contingency analysis in scenario ω

υ: Superscript/index for iteration number

Ω: Set of scenarios; index ω

$\mathcal{I}$: Set of classes

$\mathcal{I}_i$: Set of scenarios in class i

$\mathcal{S}^i$: Set of clusters for class i

$\mathcal{S}_j^i$: Set of scenarios in cluster j for class i

$\mathcal{B}$: Set of bundles

$\mathcal{B}_i$: Set of scenarios in bundle i

$|\ |$: Size of a set

Parameters :

q_i: Per MWh load shedding penalty at bus i

γ_g: Per MWh wind curtailment penalty for wind farm g

Co_g: Per MWh generation cost for generator g

ζ_l: Annual cost of line l construction

d_k: Demand at bus k

B: Diagonal matrix of line suseptance

P_g^{max}/P_g^{min}: Maximum/Minimum capacity of generator g

f_l^{max}/f_l^{min}: Maximum/Minimum capacity of line l

C^ω: Matrix of contingencies (operation states) that specifies the status of lines under different contingencies (1 for in service and 0 for out of service lines) for scenario ω; index c

$\Gamma^\omega_{m,l}$: Magnitude of violation in flow of line m when line l is on outage in scenario ω

CII^ω_l: Contingency identification index for outage of line l in scenario ω

α: Line capacity modification factor for contingency conditions ($Emergency\,capacity\,Rating = (1+\alpha)\times Normal\,capacity\,Rating$)

ϑ: Variable freezing parameter

ρ_l: Penalty factor for line l in PH algorithm

κ: Size of each bundle

$\mathcal{A}$: Clustering attributes matrix

d: Size of a TEP optimization problem

SC: Number of structural constraints for a TEP problem

CV: Number of continues variables for a TEP problem

BV: Number of binary variables for a TEP problem

Random Variables :

$\tilde{\xi^1}$: load in MW

$\tilde{\xi^2}$: Available wind output in MW

Decision Variables :

$r_{k,c}$: Load curtailment at bus k under operating state c

CW_g: Wind curtailment for wind farm g

p_g: Output power of generator g

$f_{l,c}$: Power flow in line l under operation state c

$\theta_{i,c}$: Voltage angle at bus i under operating state c. $\Delta\theta_{l,c}$ is voltage angle difference across line l under operating state c. $\Delta\theta_{l,c}$= $\theta_{k,c}$-$\theta_{n,c}$ for line l from bus k to bus n.

x_l: Binary decision variable for line l

$\boldsymbol{x}^{\omega}$: Binary decision variables vector for scenario ω

$\boldsymbol{x}^{\mathcal{B}_i}$: Binary decision variables vector for bundle $\mathcal{B}_i$

$\boldsymbol{W}_{\mathcal{B}_i}$: Multiplier vector for bundle $\mathcal{B}_i$ in PH algorithm

$\boldsymbol{\mathcal{Z}}$: Binary variables matrix for clustering

$\boldsymbol{\mathcal{H}}$: Binary variables matrix for bundling

5.1 Introduction

With increasing interest in building large-scale solar parks and wind farms and the implementation of new environmental regulations such as the "clean power plan" that will result in retirement of some conventional power plants, the need for building new transmission network is inevitable even in places in which the demand growth is not significant [96]. Planning of such network expansion is therefore increasingly important, particularly because the cost of new transmission is typically higher in real terms than historical costs.

5.1.1 A Brief Overview

The Transmission Expansion Planning (TEP) optimization problem has a long history that we briefly overview in this section. For a comprehensive overview of literature in this area, please read [57] and [49].

5.1.1.1 Solution Methods

Transmission planning methods can be divided into two main categories i.e. optimization-based, and heuristic models.

In optimization-based methods, which is the main focus of this dissertation, a mathematical formulation for TEP is developed and the problem is solved using classical optimization programming techniques. Several methods are proposed to formulate TEP problem. In [113] and [37], transmission planning is formulated as a linear optimization problem with continuous variables.

Mixed integer programming is another model that is widely used for TEP modeling ([8, 104, 84, 69] for example). A nonlinear model for TEP is developed in [119]. A complex mathematical model for centralized transmission planning and decentralized generation expansion planning is developed in [52]. Decomposition techniques like Benders decomposition [43, 94, 19, 121, 1, 54, 4], cutting-plane method [105], and Progressive Hedging [87] are also used to solve the TEP optimization problem.

In heuristic models, the TEP problem is solved through several steps of generating, evaluating, and selecting expansion plans, with or without the user's help [57]. One of the common heuristic methods is to use sensitivity analysis to select additional circuits [95, 58, 68, 82]. MISO [75], ERCOT [30], and CAISO [73] are three examples of independent system operators in the US that use different heuristic methods for TEP. As discussed in [57] and [88], existing optimization-based methods are computationally very expensive making them mainly impractical for large-scale TEP problems.

5.1.1.2 Power System Modeling

To model power flow analysis, either DC or AC models are used. Although AC models [42, 3, 100] are more accurate for power flow analysis, their nonlinear nature makes them less popular for long-term TEP problems compared to DC models [113, 8, 104, 54, 69]. Moreover, many of the parameters needed for AC analysis, such as reactive support, are not known at the time of planning the thermal transmission capacity Linear approximation of network losses, reactive power and voltage magnitude are also integrated into the DC model to improve its performance for TEP analysis [2, 120, 12]. $N-1$ contingency analysis required by NERC for power system planning and operation [91] is integrated into TEP in [104, 120, 69, 64]. Authors in [126] co-optimized transmission expansion planning with TCSC to capture the impact of FACTS devices on TEP.

5.1.1.3 Uncertainties

Fast technology changes, new policies, increasing the penetration of mobile/flexible demand along with intermittent nature of renewable resources make it hard to accurately predict future generation mix/location and demand; therefore, these uncertainties should be explicitly modeled/evaluated in TEP process by system planners. Developing a single expansion plan that considers these uncertainties using methods that heavily depend on engineering judgment can be costly and inefficient. Authors in [84], [18] evaluated the impact of ignoring uncertainties on transmission planning.

The TEP optimization problem can be formulated as a two-stage stochastic resource allocation problem (a class of mixed-integer stochastic programming) to explicitly model uncertainties using a finite set of scenarios [53]. In this formulation, in the first stage, a decision about building a new transmission line is made, and the impact of this decision on power system operation under different scenarios is evaluated in the second stage. Although formulating TEP as a two-stage stochastic optimization problem provides a strong modeling capability [94, 19, 87, 64], solving the extensive form (EF) of this problem is not tractable even for medium size problems specially when $N-1$ contingency analysis is added to the problem. Therefore, decomposition and heuristic techniques should be used for solving TEP for medium to large-scale systems.

Robust optimization is another method to integrate uncertainties into the TEP formulation. In robust optimization, uncertainties are represented using a range for each uncertain parameter instead of developing scenarios (as used by stochastic optimization), and it finds a plan that is robust for the worst case scenario. In this case, the final result is usually too conservative, which motivates an Adaptive Robust Optimization [13] formulation with budget constraint limits to mitigate the level of robustness (conservativeness of results). Authors in [105, 36, 76] formulated the TEP problem as an adaptive robust optimization. Authors in [77] applied robust optimization at distribution level decision making process. In this dissertation, we use stochastic optimization formulation to model uncertainties in TEP, corresponding to uncertainties with well-defined probability distributions.

5.1.2 Decomposition Techniques

Horizontal or *Vertical* decomposition techniques can be used to decompose a two-stage stochastic TEP optimization problem for large systems. Benders decomposition (BD) [11] is one of the widely used *horizontal* decomposition technique for solving two-stage stochastic TEP. It divides the original problem into two parts i.e master and subproblem and uses "cuts" from dual of the subproblem to model its constraints in the master problem [43]. References [43, 94, 19, 121, 1, 85, 54] applied BD to solve TEP optimization problem.

Although in several papers it is claimed that BD is easily scalable (for TEP) and can be used for real-size problems, authors in [85] showed that even for medium size networks when the number of scenarios is large (50 or more), an optimality gap between 3% to 6% would need to be accepted in the BD algorithm to get the result in a reasonable time. For large-scale problems, the subproblem itself will be hard to solve, and a large number of iterations between master and subproblem is required to meet optimality gap requirements. This problem worsens when reliability constraints are added to the TEP problem, in which subproblems should be solved for normal and under contingency operation states for all scenarios. The column-and-constraint generation method (also called cutting-plane method) is another *horizontal* decomposition technique that can be used to decompose a two-stage problem. In this method, primal "cuts" are used to represent the subproblem constraints in the master problem instead of dual cuts used by BD. Convergence guarantees and other properties of this method are explained in [51] and [7]. Authors in [105] deployed this decomposition technique for solving robust TEP.

Progressive Hedging (PH) [101] is aimed at decomposing a two-stage stochastic resource allocation problem *vertically* by solving the problem for each scenario separately, and adding *non-anticipativity* constraints to couple the first stage decision variables (standard PH). The PH method for mixed-integer problems is a heuristic method that finds an upper bound answer for the non-convex optimization problem; however, authors in [34] developed a method to calculate a lower bound for results of the PH algorithm in order to quantify the quality of results. One drawback of standard PH algorithm is that for problems with a large number of scenarios and integer variables, it may need a large number of iterations to satisfy non-anticipativity constraints (and sometimes it may never converge if no heuristic action is taken inside the algorithm). Stochastic unit commitment [108], and transmission planning [87] are examples in power system in which standard PH is applied. Authors in [24] used PH for commodity network design, and in [33], PH algorithm is used for solving multi-stage stochastic mixed integer problems.

A decision regarding the type of decomposition technique i.e *Horizontal* versus *Vertical* is usually made in advance (before problem formulation/modeling). However, depending on the size of the problem (either the network size or the number of scenarios) and the machine that is used to run the simulation, different decomposition techniques might be appropriate. For example, if the size of the network is large and a personal computer (PC) is used for simulation, probably using PH algorithm will not be a good choice because solving this problem for a single scenario by itself will be challenging, but moving from a PC to a workstation may change the situation. The same can be correct for BD when the model is developed for a problem with a small number of scenarios, and later a large number of scenarios is used to capture uncertainties. It can easily convert an efficient BD to an inefficient decomposition technique. Therefore a configurable framework is desirable.

The proposed algorithm in Chapters 2–4 significantly decreased the computational time for solving optimization-based TEP problems, but moving toward real-size problems (a reduced ERCOT case study with 3179 buses, which is 50% of the actual ERCOT network) shows that they are not scalable, and will become intractable when the size of the network significantly increases (the new case study is 10 times larger than the one used in Chapters 2–4). In this and the next chapter, a scalable and configurable decomposition framework is developed that not only provides this opportunity to use either BD or HP methods to decompose a problem but also makes it possible to use both decomposition techniques (hybrid), and takes advantages of both BD and PH for solving the same problem. Decomposing the problem by bundles of scenarios instead of each individual scenario will decrease the number of iterations in PH. But for large-scale problems, solving the extensive form (EF) of the bundled PH can be computationally expensive and even intractable. Instead of EF, we can use BD (as an efficient algorithm for problems with small-medium number of scenarios) to solve these bundled subproblems. In this way, a large-scale problem can be decomposed/parallelized both vertically and horizontally, and we can benefit from advantages of both decomposition techniques.

The rest of this chapter is organized as follows: in section 5.2, the main concept of the proposed framework explained. In subsection 5.2.1, different phases and steps of the proposed framework is explained. It is followed by a scenario bundling algorithm in subsection 5.2.2 that can be used to efficiently bundle scenarios for bundled PH algorithm. This section ends with a descriptive example that explains the developed bundling algorithms for a simple case study. In section 5.3, mathematical formulation related to different parts of the developed framework is provided. In subsection 5.3.1, the stochastic TEP formulation with security constraints (similar to Chapter 4) is provided. It is followed by bundled PH algorithm and formulation in subsection 5.3.2. The integer programming formulation for clustering is explained in subsection 5.3.3. This chapter is finalized with a review of the VCL algorithm from Chapter 2 that is used for creating bundling attributes. The performance of the proposed framework and numerical results are given in Chapter 6.

5.2 The Proposed Framework

5.2.1 Framework Overview

The proposed framework is designed to be flexible and configurable for different problem sizes on different machines. It can be configured to solve a problem in extensive form (EF), or using PH, BD, and Hybrid techniques that provides more flexibility from the modeling perspective. The proposed framework can be summarized as follows:

Phase 0: Data preparation

Step 1: Input data and set parameters

Input data includes base network, scenarios, and candidate lines list. In this step, the planner configures the framework by setting parameters; i.e. the number of scenarios in each bundle (κ) and the type of decomposition technique that should be used (PH, BD or Hybrid) for phases I and II. Settings for phase II can be modified later in step 4 if it is necessary.

Phase I: TEP without contingency analysis

Step 2: Scenario bundling

In this step, OPF for the base (existing) network is solved, and calculated load shedding and wind curtailment will be used to develop an attribute for scenario bundling. After developing appropriate criteria, scenarios are distributed between groups using the developed scenario bundling method (see subsection 5.2.2).

Step 3: Solving TEP

In this step, based on inputs from step 1 and bundles from step 2, TEP for normal operation states is solved. This step can be parallelized. The proposed method in Chapter 3 can be used to solve TEP in this step faster.

Phase II: TEP with contingency analysis

This phase is run if contingency analysis should be integrated in TEP process.

Step 4: Scenario Bundling

Based on parameter settings, the scenario bundling method (see subsection 5.2.2 for more detail) is used to bundle scenarios. The VCL algorithm developed in Chapter 2 is used to develop bundling criteria for this step.

Step 5: Solving TEP with contingency analysis

In this step, TEP with contingency analysis is solved. Based on framework's setting, either PH, BD, or hybrid may be used for solving this large-scale optimization problem. This step can be parallelized if PH and/or BD are selected as the solving algorithm. The developed algorithm in Chapter 4 can be used for solving TEP for each subproblem in this step.

Phase III: Quantifying the quality of results

Step 6: Calculating a lower bound answer

If PH or hybrid is selected for phase I and/or II, then finding a lower bound answer is necessary to quantify the quality of results. In this step, the proposed lower bound formulation for PH in [34] is used to calculate a lower bound.

Step 7: Calculate optimality gap

The optimality gap (ε) can be calculated using the upper bound from step 5 (or step 3 in case of TEP without contingency analysis) and the lower bound from step 6. The selected plan is $\varepsilon -$

suboptimal.

The proposed framework is summarized in the flowchart in Figure 5.1.

5.2.2 Developed Scenario Bundling Method

In this chapter, a heuristic method is developed to bundle scenarios. The main purpose of this method is to create heterogeneous groups of scenarios with minimum dissimilarity *between* the groups collectively (based on selected attributes/criteria) and with relatively the same computational burden. Having similar bundles will improve the performance of PH algorithm by facilitating convergence of non-anticipativity constraints, as for a set of identical groups of scenarios, PH only needs one iteration to converge (although the choice of bundling does not necessarily reduce computational time). In contrast with clustering in which the objective is to minimize dissimilarity *within* groups (by forming homogeneous groups), scenario bundling tries to minimize dissimilarity *between* groups (mathematical formulation is provided in section 5.3.4). Developed groups partition the scenarios, and their size (κ) is constant for each phase. The proposed method bundles scenarios through three steps i.e. classification, clustering, and grouping. These steps are explained in the following subsections. It should be noted that scenario bundling is required only if $1 < \kappa < |\Omega|$, where Ω is the set of all scenarios and $|\Omega|$ represents the size of this set.

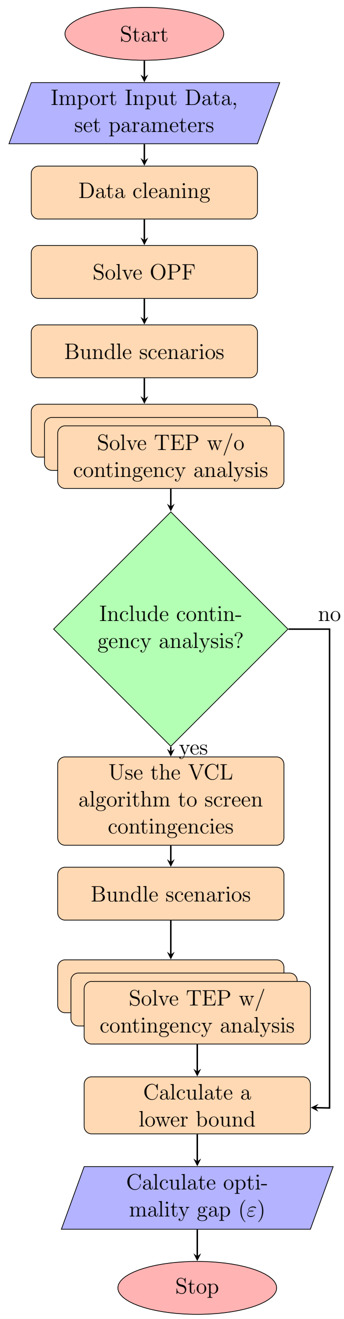


Figure 5.1: Flowchart of the proposed generalized framework

5.2.2.1 Classification

In classification, a model or classifier is constructed to predict class labels such as "safe" or "risky" for bank loan application, or "light" and "heavy" loading conditions for electric networks. There are different classification methods such as Decision Tree Induction, Bayes Classification methods, and Rule-Based classification [47]. We use the Rule-Based method, because its structure allows us to easily integrate expert knowledge into the bundling process. It has the following structure:

$$\textbf{IF} \quad Condition \quad \textbf{THEN} \quad Conclusion \tag{5.1}$$

For our banking example, it can be written as

$$\textbf{IF} \quad age \leq 25 \; AND \; student \quad \textbf{THEN} \quad Safe$$

For electric network example, we can have

$$\textbf{IF} \; average\, line\, loading \geq 50\% \; \textbf{THEN} \; Heavily\, loaded$$

Rule-based classification will partition the original scenario set Ω into a finite number of non-empty classes $\mathcal{I} = \{\mathcal{I}_1, \ldots, \mathcal{I}_q\}$.

Different classification rules can be defined depending on the purpose of a study. For numerical analysis in section 6.3, we will use the number of important lines for contingency analysis (ICLs) as a classifier in step 4. By using this classifier we may need to adjust the number of scenarios in classes (those that are close to boundaries) for feasibility of clustering. Classification is an optional part of the bundling process, and if there is no classifier, then there will be only one class that includes all scenarios ($\mathcal{I} = \{\mathcal{I}_1\}$).

5.2.2.2 Clustering

Clustering is the process of grouping a set of objects in a way that objects within a cluster have the highest similarity. In this step, similar scenarios in each class ($\mathcal{I}_i$) are clustered based on selected attribute/developed criteria, and form the set $\mathcal{S}^i = \{\mathcal{S}^i_1, \ldots, \mathcal{S}^i_c\}$. Without loss of generality, scenarios are clustered in groups with the same size in this chapter, and the size of each cluster ($\mathcal{C}_s$) can be calculated from the following equation.

$$\mathcal{C}_s = \frac{|\Omega|}{\kappa} \tag{5.2}$$

where, we assume that $|\Omega|$ is dividable by κ.

It is important to choose an attribute/criteria that is appropriate for the purpose of the study and provides insight for grouping phase. For example, bundles from step 2 of the proposed framework are used for solving TEP in step 3. Load shedding and wind curtailment (under normal operation condition) are highly penalized (compared to generation operation cost) in the TEP objective function (4.5); therefore, load (and wind) will be curtailed only if there is not enough transmission capacity to supply them (and transfer their output), which is a signal for a need for transmission expansion. Therefore, we used these two components in the objective function to form a clustering attribute for phase I. A weighted sum of load shedding and wind curtailment

(LW) is defined as a clustering attribute for this step. For step 4 of the framework, selected lines for contingency analysis for each scenario (ICL^{ω}s) is used as an attribute for scenario clustering because TEP with contingency analysis is solved in step 5, and ICL^{ω}s can significantly affect the selected expansion plan [69].

Partitioning method is used to create clusters by minimizing distance between different attributes of objects (scenarios here). For step 2, scenarios with closest LW values are clustered together, and for step 4, the objective of clustering optimization problem is to maximize similarity of ICL^{ω}s within each cluster. It creates a good "base" for grouping phase. An integer programming problem is solved to cluster scenarios in steps 2 and 4 (see section 5.3.3 for mathematical formulation).

5.2.2.3 Grouping into Bundles

In the last phase of the proposed scenario bundling method, members of each cluster are distributed between groups (bundles) with the objective of minimizing dissimilarity *between* groups (by forming heterogeneous bundles). For the scenario set Ω, a bundle set $\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_b\}$ of non-empty and mutually exclusive subsets ($\forall i \neq j,\ \mathcal{B}_i \bigcap \mathcal{B}_j = \emptyset$ and $\bigcup_j \mathcal{B}_j = \Omega$) is formed. As scenarios in each cluster share some similar characteristics, one way is to distribute members of each cluster randomly between groups. It is also possible to define new criteria for grouping in this step. For developing new criteria, two main points should be noticed: first, the criteria should be at the group level rather than the scenario level because increasing similarity *between* groups is the purpose of this step. Second, the new criteria should not be significantly different compared to classification/clustering criteria, because the implicit assumption in this step is that scenarios in each cluster share similar attributes, and this assumption is mainly valid for the attributes used in classification and clustering phases. Ignoring these points may decrease similarity between formed groups.

For step 2, scenarios are distributed between groups so that groups have relatively the same aggregated LW value ($LW_{\mathcal{B}_i}$) because of its major contribution in the objective function in TEP optimization problem in step 3. For step 4, total number of ICLs in each group is used as a criteria for distributing scenarios between bundling groups. This attribute will result in forming groups with relatively the same number of operation states, which will have a huge impact on computational time. Combining this criteria with the one used for clustering will result in creating groups that have relatively the same impact on optimal result (because of similar lines for contingency analysis) and requires relatively the same computational burden (number of operation states).

As a separate stochastic TEP is solved for each bundle in PH algorithm, the probability of each scenario should be updated based on equations (5.3) and (5.4):

$$P_{\mathcal{B}_i} = \sum_{\omega \in \mathcal{B}_i} P^{\omega} \quad \forall \mathcal{B}_i \in \mathcal{B} \tag{5.3}$$

$$Pu^{\omega} = \frac{P^{\omega}}{P_{\mathcal{B}_i}} \quad \forall \omega \in \mathcal{B}_i, \forall \mathcal{B}_i \in \mathcal{B} \tag{5.4}$$

$$|\Omega| = \sum_{\mathcal{B}_i \in \mathcal{B}} |\mathcal{B}_i| \tag{5.5}$$

$$\sum_{\mathcal{B}_i \in \mathcal{B}} P_{\mathcal{B}_i} = 1 \tag{5.6}$$

where, P^{ω} is the original probability of scenario ω, $P_{\mathcal{B}_i}$ is probability of bundle $\mathcal{B}_i$ in set of bundles $\mathcal{B}$, and Pu^{ω} is updated probability of scenario ω as a member of bundle $\mathcal{B}_i$. Equations (5.5) and (5.6) enforce scenario bundling to be mutually exclusive.

Authors in [108] suggested that forming bundles with two scenarios may improve the performance of the PH algorithm for stochastic unit commitment problem, but they did not discuss how bundles should be formed. In [24], authors proposed a scenario grouping method for commodity transportation network planning, in which the objective of grouping is to maximize dissimilarity within groups (replacing minimization in equation (5.28) with maximization). Compared to [24], the proposed method in this paper minimizes dissimilarity between groups (using the objective function (5.37)), take into account the existing hardware infrastructure to control the size of each bundle, and forms bundles with relatively the same size to improve the performance of parallelizing (see section 6.2.7 for more details).

5.2.3 A Descriptive Example

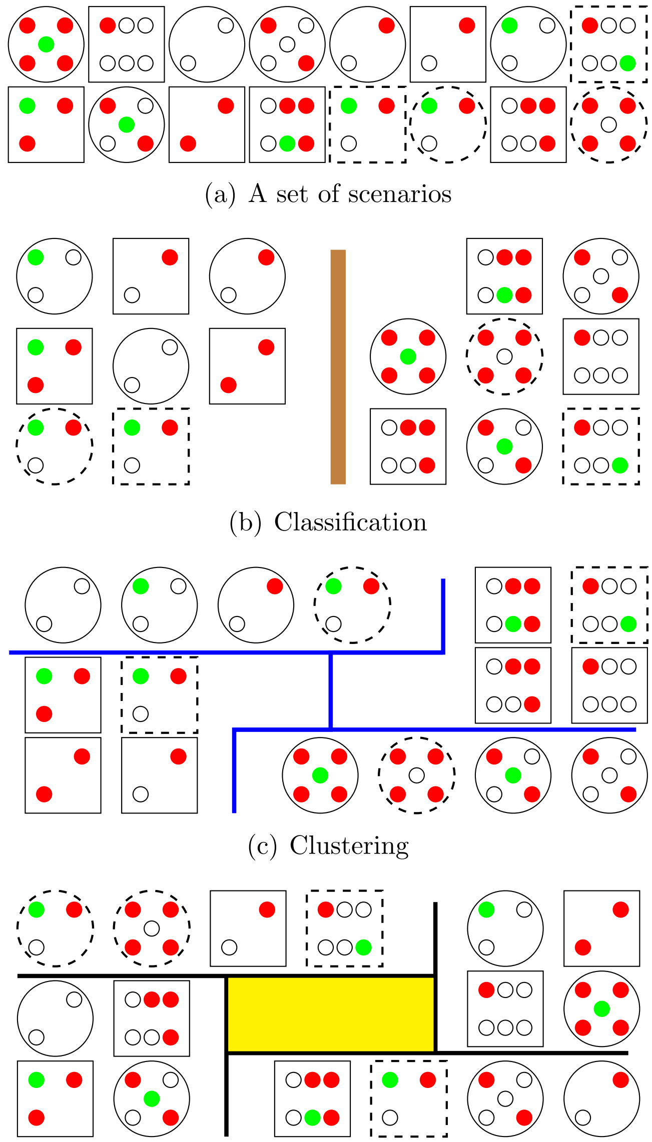
In this section, a descriptive example is used to explain implementation of all steps of the developed scenario bundling method (from section 5.2.2). Figure 5.2(a) shows a set of 16 scenarios ($|\Omega| = 16$) that are developed to capture uncertainties in wind, load, and future marker regulation on CO_2 emission for planning purposes. The target is to bundle scenarios into groups of 4 ($\kappa = 4$) with the objective function of minimizing the dissimilarity between groups. Shape (rectangular for high load and circle for high wind uncertainties), design (dashed lines represent the future market with CO_2 penalty, and solid lines for a future without CO_2 penalty), number of dots (shows the number of overloaded lines in the base case), and colors (the level of overload in lines) are used to visualize different attributes of scenarios. In the first step, scenarios are classified based on the number overloaded lines in the scenarios using the following rule:

$$\textbf{IF } number\, of\, overloaded\, lines \geq 5$$
$$\textbf{THEN } Heavly\, loaded\, network$$

In Figure 5.2(b), the vertical brown line separates scenarios into two classes ($\mathcal{I} = \{\mathcal{I}_1, \mathcal{I}_2\}$) based on their impact on network loading.

In the next step, scenarios in each class $\mathcal{I}_1$ and $\mathcal{I}_2$ are clustered based on similarity in uncertainties that they represent (their shapes). Based on (5.2), the size of clusters is equal to 4 ($\mathcal{C}_s = 4$). As the number of scenarios in each class is 8 ($|\mathcal{I}_1| = |\mathcal{I}_2| = 8$) and clustering with $\mathcal{C}_s = 4$ is feasible for each class, we do not need to modify the size of classes for this case. In Figure 5.2(c), clusters are separated with blue lines ($\mathcal{S}^1 = \{\mathcal{S}^1_1, \mathcal{S}^1_2\}, \mathcal{S}^2 = \{\mathcal{S}^2_1, \mathcal{S}^2_2\}$). In the last step, scenarios in clusters are distributed between our 4 target groups ($|\mathcal{B}| = 4$) with the criteria that the number of overloaded lines and their level of overload (color here) have the most similarity (shown in Figure 5.2(d)), forming $\mathcal{B} = \{\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3, \mathcal{B}_4\}$. Now we have 4 groups of scenarios, each including 2 high load related scenarios and 2 high wind related scenarios with 16 overloaded lines in each (7 at white level, 7 at red level, and 2 at green level).

It should be noted that similarity between these groups is only valid for attributes used in the bundling process. For example, the impact of having or not having CO_2 penalty (dashed versus solid lines) is considered as an attribute in neither of three bundling steps, and results (Figure 5.2(d)) show that there is no similarity *between* groups for this attribute. These bundles of scenarios may not improve the performance of bundled PH algorithm if CO_2 penalty significantly affects the selected lines for transmission expansion (for example in systems with high penetration of cheap coal power plants).



(a) A set of scenarios

(b) Classification

(c) Clustering

(d) Grouping

Figure 5.2: An example for explaining different steps of the bundling method

5.3 Problem Formulation

5.3.1 Two-Stage stochastic TEP Formulation

As discused in Chapter 4, stochastic programming is one of the widely used methods to model uncertainties (by developing different scenarios) in decision making process for resource allocation problems. Uncertainties in long-term transmission expansion can be categorized as macro uncertainties such as changes in market rules, environmental constraints or new technologies, and micro uncertainties such as hourly wind/solar and load variations [64]. To capture these uncertainties, different scenario generation/reduction methods might be used. The quality of scenarios is critical and can significantly affect the selected expansion plan. For example, in ERCOT, historical data along with workshops with stakeholders are used to develop scenarios for long-term TEP [29]. For a given scenario set, we develop a framework to efficiently solve this optimization problem; therefore, the nature of uncertainty and the origin of scenarios is not our concern in this work. It should be mentioned that minimizing the expected value is a better criterion for micro uncertainties. The two-stage stochastic TEP is formulated as follows:

$$Z^{*} = \min \zeta^{\intercal} \boldsymbol{x} + \mathbb{E}[Q(\boldsymbol{x}, \tilde{\xi})] \tag{5.7}$$

$$\text{st.} \quad \boldsymbol{x} \in \{0,1\}^{|N_l|} \tag{5.8}$$

$\mathbb{E}[Q(\boldsymbol{x}, \tilde{\xi})]$ represents the expected value of operation costs including load shedding and wind curtailment penalty and generation costs for TEP problem formulation. This expected value is approximated with a weighted sum of a limited number of scenarios as follows [31]:

$$\mathbb{E}[Q(\boldsymbol{x},\tilde{\xi})] \approx \sum_{\Omega} P^{\omega} Q(\boldsymbol{x},\xi^{\omega}) \tag{5.9}$$

where $Q(\boldsymbol{x},\xi)$ is the optimal value of power system operation for a given scenario ω [64].

$$Q(\boldsymbol{x},\xi) = \min \sum_{N_s} (\sum_{N_b} q_k r_{k,c}) + \sum_{N_w g} \gamma_g CW_g + \sum_{Ng} Co_g p_g \tag{5.10}$$

$$\text{st.} \quad -\sum_{L_k} f_{l,c} + \sum_{G_k} p_g + r_{k,c} = d_k \tag{5.11}$$

$$-\mathcal{M}_l(1 - C_{l,c}x_l) \leq f_{l,c} - B_{l,l}\Delta\theta_{l,c} \tag{5.12}$$

$$\mathcal{M}_l(1 - C_{l,c}x_l) \geq f_{l,c} - B_{l,l}\Delta\theta_{l,c} \tag{5.13}$$

$$CW_g \geq (P_g^{max} - p_g) \tag{5.14}$$

$$(C_{l,c}x_l) f_l^{min} \leq f_{l,c} \leq f_l^{max}(C_{l,c}x_l) \tag{5.15}$$

$$P_g^{min} \leq p_g \leq P_g^{max} \tag{5.16}$$

$$0 \leq r_{k,c} \leq d_k \tag{5.17}$$

$$-\frac{\pi}{2} \leq \theta_{k,c} \leq \frac{\pi}{2} \tag{5.18}$$

$$CW_g \geq 0 \tag{5.19}$$

$$x_l = 1, \quad \forall l \in N_o \tag{5.20}$$

$$x_l \in \{0,1\}, \quad \forall l \in N_l \tag{5.21}$$

In equation (5.10), load shedding is penalized over all operating states (N_s) to satisfy the $N-1$ criterion (no load shedding is accepted during both normal and under single contingency states). Equation (5.11) enforces power balance at each bus. Equations (5.12) and (5.13) show DC representation of flow in transmission lines with big-$\mathcal{M}$ technique. Equation (5.14) measures wind curtailment at each bus. Equation (5.15) shows flow in all lines should always be between their maximum and minimum capacity limits. These limits will be modified based on the given value for α for emergency conditions (contingency in the network). Equations (5.16)-(5.18) enforce power plants' dispatch, load shedding and voltage angles, respectively, to be between their minimum and maximum limits. Equation (5.19) enforces non-negativity of wind curtailment. Equation (5.20) sets decision variables for existing lines to 1. Equation (5.21) enforces that x_l is a binary decision variable for transmission lines ($x_l = 1$ when line l is built and $x_l = 0$ when line l is not built).

Constraints (5.11)–(5.19) represents lossless DC power flow model. Authors in [37] and [9] showed that DC model is accurate enough for long-term planning purpose because of the large level of simplifications in other aspects, but network losses, reactive power and voltage magnitude might be critical in some networks. Authors in [12, 120, 2, 21, 22] provide models to improve the performance of DC model by adding linear approximation of reactive power, network losses, and voltage magnitudes. As these models all preserve linearity of power flow model, they can be added to the model in this chapter.

Depending on the size of the network and the number of scenarios, solving the extensive form of problem (5.7) can be extremely computationally expensive. Therefore, decomposition techniques are used to find a near-optimal answer for large-scale problems. In the next section, PH with bundled scenarios is explained as the base for steps 3 and 5 in our proposed framework. Details on BD technique is not in the scope of this chapter and can be found in [23].

5.3.2 Progressive Hedging Algorithm with Bundled Scenarios

Progressive Hedging [101] is one of the decomposition techniques that can be used for solving two-stage (or multi-stage) stochastic mixed integer optimization problems. The standard PH algorithm separates the problem vertically, and solves it for each scenario individually. The TEP problem (5.7) can be rewritten as the following so-called *scenario* formulation:

$$Z^{*}=\min \sum_{\Omega} P^{\omega}[\zeta^{\intercal}\boldsymbol{x}^{\omega}+Q(\boldsymbol{x}^{\omega},\xi^{\omega})] \tag{5.22}$$

$$\text{st.} \quad \boldsymbol{x} \in \{0,1\}^{|N_l|} \tag{5.23}$$

$$\boldsymbol{x}^1 = \cdots = \boldsymbol{x}^s \tag{5.24}$$

A copy of decision variable vector $\boldsymbol{x}^{\omega}$ is created for each scenario ω in Ω that allows solution of the TEP problem for each scenario independently, and non-anticipativity constraints (5.24) are added to couple first stage solutions and guarantee that the final expansion plan does not depend on scenarios.

Instead of decomposing the problem for each individual scenario, it is possible to use bundles of scenarios ($\mathcal{B} = \{\mathcal{B}_1, \ldots, \mathcal{B}_b\}$) for decomposition as discussed in sections5.2.2 and 5.2.3. Equations (5.22)–(5.24) can be rewritten for bundled PH as follows:

$$Z^{*}=\min\sum_{\mathcal{B}}[P_{\mathcal{B}_i}(\zeta^{\intercal}\boldsymbol{x}^{\mathcal{B}_i})+\sum_{\mathcal{B}_i}Pu^{\omega}Q(\boldsymbol{x}^{\mathcal{B}_i},\xi^{\omega})] \tag{5.25}$$

$$\text{st.}\quad \boldsymbol{x}\in\{0,1\}^{|N_l|} \tag{5.26}$$

$$\boldsymbol{x}^{\mathcal{B}_1}=\cdots=\boldsymbol{x}^{\mathcal{B}_b} \tag{5.27}$$

In this case, a copy of decision variable vector $\boldsymbol{x}^{\mathcal{B}_i}$ is created for all $\mathcal{B}_i$s in $\mathcal{B}$. Non-anticipativity constraints (5.27) are explicitly modeled for scenario bundles, and they are implicitly modeled for scenarios within each bundle (κ scenarios in each bundle already have the same first stage decision variable $\boldsymbol{x}^{\mathcal{B}_i}$) that usually reduces the number of iterations compared to standard PH.

Initialization: $\upsilon \leftarrow 1, \boldsymbol{W}^{\upsilon}_{\mathcal{B}_i} \leftarrow \boldsymbol{0}\ \forall \mathcal{B}_i \in \mathcal{B}$
for $\forall \mathcal{B}_i \in \mathcal{B}$ **do**
    $\boldsymbol{x}^{\mathcal{B}_i,\upsilon} \leftarrow argmin\, \boldsymbol{\zeta}^{\intercal}\boldsymbol{x}^{\mathcal{B}_i} + \sum_{\omega\in\mathcal{B}_i} Pu^{\omega}Q(\boldsymbol{x}^{\mathcal{B}_i},\boldsymbol{\xi}^{\omega})$
end for
Aggregation: $\hat{\boldsymbol{x}}^{\upsilon} \leftarrow \sum_{\mathcal{B}} P_{\mathcal{B}_i}\boldsymbol{x}^{\mathcal{B}_i,\upsilon}$
$Err \leftarrow \sum P_{\mathcal{B}_i}\|\boldsymbol{x}^{\mathcal{B}_i,\upsilon} - \hat{\boldsymbol{x}}^{\upsilon}\|$
while $Err \geq \epsilon$ **do**
    $\upsilon \leftarrow \upsilon + 1$
    $\boldsymbol{W}^{\upsilon}_{\mathcal{B}_i} \leftarrow \boldsymbol{W}^{\upsilon-1}_{\mathcal{B}_i} + \boldsymbol{\rho}^{\intercal}(\boldsymbol{x}^{\mathcal{B}_i,\upsilon-1} - \hat{\boldsymbol{x}}^{\upsilon-1})$
    for $\forall \mathcal{B}_i \in \mathcal{B}$ **do**
        $\boldsymbol{x}^{\mathcal{B}_i,\upsilon} \leftarrow argmin\, \boldsymbol{\zeta}^{\intercal}\boldsymbol{x}^{\mathcal{B}_i} + \sum_{\omega\in\mathcal{B}_i} Pu^{\omega}Q(\boldsymbol{x}^{\mathcal{B}_i},\boldsymbol{\xi}^{\omega}) + \boldsymbol{W}^{\upsilon}_{\mathcal{B}_i}{}^{\intercal}\boldsymbol{x}^{\mathcal{B}_i} + \frac{\boldsymbol{\rho}^{\intercal}}{2}(\boldsymbol{x}^{\mathcal{B}_i} - \hat{\boldsymbol{x}}^{\upsilon-1})^2$
    end for
    Aggregation: $\hat{\boldsymbol{x}}^{\upsilon} \leftarrow \sum_{\mathcal{B}} P_{\mathcal{B}_i}\boldsymbol{x}^{\mathcal{B}_i,\upsilon}$
    $Err \leftarrow \sum P_{\mathcal{B}_i}\|\boldsymbol{x}^{\mathcal{B}_i,\upsilon} - \hat{\boldsymbol{x}}^{\upsilon}\|$
end while

Figure 5.3: Progressive Hedging Algorithm with bundled scenarios

Through an iterative process, PH will converge to a unique answer for the first stage decision variables by penalizing deviations of non-anticipative variables from their mean values. The PH algorithm with bundled scenarios is shown in Figure 5.3. In the first line, the initial value of the iteration counter (υ), and multiplier vector ($\boldsymbol{W}^{\upsilon}_{\mathcal{B}_i}$) is set. From line 2–4, the TEP optimization problem for each bundle is solved separately (and can be parallelized). The proposed algorithm in Chapter 3 can be used for this step. In line 5, the weighted sum of individual expansion plans ($\boldsymbol{x}^{\mathcal{B}_i,\upsilon}$s) is calculated. Line 6 calculates the deviation (Err) from averaged expansion plan ($\hat{\boldsymbol{x}}^{\upsilon}$). Lines 7–15 cover the main iterative part of the bundled PH algorithm. In line 8, the value of counter is updated. Line 9 updates the value of multiplier vector by using penalty vector $\boldsymbol{\rho}$. Lines 10–12 solve an updated TEP formulation with multiplier and penalizing deviation from average value of first stage decision variables. This optimization problem is solved for each bundle independently, so they can be solved in parallel. As each subproblem will be less computationally expensive, we can benefit from the proposed algorithm in Chapter 4 for this purpose. Lines 13 and 14 update the calculated average value for $\boldsymbol{x}$ and Err, respectively.

5.3.3 Clustering Algorithm

As defined in [47], "cluster analysis or clustering is the process of partitioning a set of data objects (or observations) into subsets. Each subset is a cluster such that objects in a cluster are similar to one another, yet dissimilar to objects in other clusters." Major fundamental clustering methods can be classified into four categories i.e. Partitioning methods, Hierarchical methods, Density-based methods, and Grid-based methods. A detailed discussion on each category can be found in [47].

A partitioning method can be used to find mutually exclusive clusters based on distances between its elements. For a finite set (for example, $\mathcal{I}_1$), a set $\mathcal{S}^1 = \{\mathcal{S}^1_1, \ldots, \mathcal{S}^1_c\}$ of non-empty subsets of $\mathcal{S}^1_i$ is a partition if $\forall k \neq j,\ \mathcal{S}^1_k \bigcap \mathcal{S}^1_j = \emptyset$ and $\bigcup_j \mathcal{S}^1_j = \mathcal{I}_1$. Partitioning can be formulated as an integer programming problem in which the objective is to minimize the distance (Euclidean distance here) between members of each cluster based on selected attribute(s).

$$IP = \min \sum_{m=1}^{n_a} \sum_{k=1}^{n_c} \sum_{i=1}^{n_s} \sum_{j=1}^{n_s} \|\mathcal{A}_{i,m}\mathcal{Z}_{i,k} - \mathcal{A}_{j,m}\mathcal{Z}_{j,k}\|^2 \tag{5.28}$$

$$\text{st.} \sum_{k=1}^{n_c} \mathcal{Z}_{i,k} = 1, \quad \forall i \in \mathcal{I}_1 \tag{5.29}$$

$$\sum_{i=1}^{n_s} \mathcal{Z}_{i,k} = \mathcal{C}_s \quad \forall k \in \mathcal{S}^1 \tag{5.30}$$

where $\mathcal{A}$ is clustering attribute matrix ($[n_s \times n_a]$) for set $\mathcal{I}_1$, n_c is the number of clusters ($n_c = |\mathcal{S}^1|$), n_s is the number of scenarios in set $\mathcal{I}_1$ ($n_s = |\mathcal{I}_1|$), n_a is the number of attributes, $\mathcal{C}_s$ is the number of scenarios in each cluster ($\mathcal{C}_s = \frac{n_s}{n_c}$, equivalent to (5.2)), and $\mathcal{Z}$ is the binary decision variables matrix

($[n_s \times n_c]$) that assigns scenarios to clusters.

$$\mathcal{Z} = \begin{bmatrix} \mathcal{Z}_{1,1} & \cdots & \mathcal{Z}_{1,n_c} \\ \vdots & \ddots & \vdots \\ \mathcal{Z}_{n_s,1} & \cdots & \mathcal{Z}_{n_s,n_c} \end{bmatrix}$$

Equation (5.29) enforces that each scenario can only be a member of one cluster. Equation (5.30) enforces that all scenarios should be assigned to clusters and the size of all clusters is equal to $\mathcal{C}_s$. This is designed based on the assumption that we made in this chapter. However, to have a flexible cluster size, equation (5.30) can be replaced with (5.31) and (5.32):

$$\sum_{k=1}^{n_c} \sum_{i=1}^{n_s} \mathcal{Z}_{i,k} = n_s \tag{5.31}$$

$$\sum_{i=1}^{n_s} \mathcal{Z}_{i,k} \geq 1 \quad \forall k \in \mathcal{S}^1 \tag{5.32}$$

Equation (5.32) guarantees that there will be no empty cluster.

The objective function (5.28) is nonlinear. As $\mathcal{Z}$ is a matrix of binary decision variables, for all $i = j$ the nonlinear term $\mathcal{Z}_{i,k}\mathcal{Z}_{j,k}$ can be replaced with $\mathcal{Z}_{i,k}$. For $i \neq j$, the nonlinear term $\mathcal{Z}_{i,k}\mathcal{Z}_{j,k}$ can be replaced with a new binary variable $\mathcal{Y}_r$, and constraints (5.33)-(5.35) should be added to the IP problem:

$$\mathcal{Y}_r \leq \mathcal{Z}_{i,k} \tag{5.33}$$

$$\mathcal{Y}_r \leq \mathcal{Z}_{j,k} \tag{5.34}$$

$$\mathcal{Y}_r \geq \mathcal{Z}_{i,k} + \mathcal{Z}_{j,k} - 1 \tag{5.35}$$

The size of matrix $\boldsymbol{\mathcal{Y}}$ ($[n_r \times n_c]$) for a scenario set of n_s scenarios is equal to:

$$n_r = \frac{n_s \times (n_s - 1)}{2} \tag{5.36}$$

It should be mentioned that for cases in which a very large number of scenarios should be clustered, solving IP can be computationally expensive. There are heuristics such as k-means methods that can be used for partitioning. Details of these methods are not in the scope of this paper and can be found in [47].

5.3.4 Scenario Bundling Algorithm

As stated in section 5.2.2, the main goal of scenario bundling is to maximize similarity (minimizing dissimilarity) *between* bundles to improve the performance of bundled PH algorithm. This problem can be formulated as an integer programming problem. The mathematical formulation for scenario bundling is as follows.

$$\min \sum_{m=1}^{n_a} \sum_{b,b'=1}^{n_b} \| \mathcal{Q}_{b,m} - \mathcal{Q}_{b',m} \|^2 \tag{5.37}$$

$$\text{st. } \mathcal{Q}_{b,m} = \texttt{mean}(\sum_{i=1}^{n_s} \mathcal{A}_{i,m} \mathcal{H}_{i,b}), \ \forall b \in \mathcal{B}, \forall m \in \mathcal{A} \tag{5.38}$$

$$\sum_{b=1}^{n_b} \mathcal{H}_{i,b} = 1, \quad \forall i \in \Omega \tag{5.39}$$

$$\sum_{i=1}^{n_s} \mathcal{H}_{i,b} = \kappa \quad \forall b \in \mathcal{B} \tag{5.40}$$

Where $\mathcal{Q}_{b,m}$ is the average value of attribute m in bundle b that can be calculated from (5.38), and $\boldsymbol{\mathcal{H}}$ is the binary decision variable matrix for bundling. The objective function (5.37) maximizes the similarity between bundles by minimizing the distance between `mean` value of attributes of bundles. Equation (5.39) enforces each scenario should be assigned to a bundle, and equation (5.40) enforces the size of each bundle.

Solving this problem for a large set of scenarios can be computationally expensive; therefore, a heuristic method is developed in section 5.2.2 to solve this problem faster.

5.3.5 Variable Contingency List (VCL) Algorithm

Based on Chapter 2, modified Line Outage Distribution Factors (LODFs) are used to estimate post-contingency flow in transmission lines when one line is on outage. The following equations are used to create important contingency lists for different scenarios:

$$\Gamma_{m,l}^{\omega} = \frac{f_{m,l}^{\omega} - f_{m}^{max}}{f_{m}^{max}} \qquad , \forall m, l \in N_o, \forall \omega \in \Omega \tag{5.41}$$

$$\Phi_l^{\omega} = \{ m \in N_o \,|\Gamma_{m,l}^{\omega} \geq \alpha \} \quad , \forall l \in N_o, \forall \omega \in \Omega \tag{5.42}$$

$$CII_l^{\omega} = \begin{cases} \frac{\sum_{\Phi_l^{\omega}} \Gamma_{m,l}^{\omega}}{|\Phi_l^{\omega}|}, & if \quad |\Phi_l^{\omega}| \neq 0 \\ 0, & if \quad |\Phi_l^{\omega}| = 0 \end{cases} \tag{5.43}$$

$$ICL^{\omega} = \{ l \in N_o \,|CII_l^{\omega} \geq \alpha \} \quad , \forall \omega \in \Omega, \tag{5.44}$$

where (5.41) calculates over/under loading on line m when line l is out. In this equation, $f_{m,l}^{\omega}$ represents the magnitude of post-contingency flow in line

m when line l is on outage. Equations (5.42)-(5.43) are used to calculate Contingency Identification Index (CII) for each scenario with α as the line capacity modification factor during contingencies that accounts for emergency or short-term rating of lines. Equation (5.44) creates important contingency list (ICL) based on CII (see Chapters 2 and 4 for more details).

A discussion on the performance of the proposed framework and numerical results are provided in the next chapter.

Chapter 6

Decomposition Framework: Model Performance and Numerical Results[1]

6.1 Introduction

In Chapter 5, a configurable framework is proposed to solve different problems. In this chapter, we investigate its performance from several perspectives i.e. parameter setting, choice of a decomposition algorithm, linking PH and BD, PH performance improvement, optimality gap, parallelizing, scalability and maintainability in section 6.2. In section 6.3, the proposed method is applied to two case studies i.e. a 13-bus system with 100 scenarios and a reduced ERCOT system with 3179 buses, 4458 branches and 10 scenarios. For each case study, the results are compared with standard PH, randomly bundled PH, and developed method in Chapter 4 to evaluate different aspects of the proposed framework.

[1]Mohammad Majidi-Qadikolai and Ross Baldick. A generalized decomposition framework for large-scale transmission expansion planning. IEEE Transactions on Power Systems, PP(99):1-1, 2017. Authors had equal contributions.

Table 6.1: Framework performance under different settings

	PH	BD	Hybrid
$\kappa = 1$	PH	Heuristic	Hybrid
$1 < \kappa < \lvert\Omega\rvert$	PH	Heuristic	Hybrid
$\kappa = \lvert\Omega\rvert$	EF	BD	BD

6.2 Model Performance Discussion

6.2.1 Parameter settings for the framework

The size of each bundle (κ) and the choice of a decomposition method are set in step 1 in the framework (see section 5.2.1). Table 6.1 shows different possible combinations for setting these two parameters. For the PH algorithm, by setting $\kappa = 1$ a standard PH is solved, $1 < \kappa < |\Omega|$ will result in a bundled PH, and $\kappa = |\Omega|$ is equivalent to solving the extensive form (EF) of the optimization problem. If BD is selected as the solving method, then for $1 \leq \kappa < |\Omega|$, the problem is solved separately for each bundle, and a heuristic method should be used to select a unique first stage answer. For $\kappa = |\Omega|$, a standard BD is solved. When Hybrid method is selected, for $1 \leq \kappa < |\Omega|$, both PH and BD are used for solving the problem in steps 3 and/or 5 in the framework. This is discussed more in section 6.2.3. For $\kappa = |\Omega|$, hybrid method will be the same as BD method. It should be mentioned that these parameters can be set independently for phases I and II providing more flexibility, potentially improving the effectiveness of the proposed framework.

6.2.2 Factors affecting the choice of parameters

The size of the problem, the design of decomposition algorithms, existing hardware infrastructure, and solvers are critical for making a decision about setting parameters for the proposed framework. We briefly overview these factors in the following.

6.2.2.1 The size of the problem (d)

The number of structural constraints (SC), equations (4.6)-(4.9), continuous (CV) and binary (BV) decision variables are main factors for the size of the TEP optimization problem. For the extensive form of this TEP formulation from section **??** (depending on the choice and design of decomposition algorithms, new variables and constraints may be added), these values can be calculated from the following equations:

$$d = \{SC, CV, BV\} \tag{6.1}$$

$$SC = (2 \times (|N_b| + |N_l|) \times |N_s^{\omega}| + |N_{wg}|) \times |\Omega| \tag{6.2}$$

$$CV = ((2 \times |N_b| + |N_l|) \times |N_s^{\omega}| + |N_g| + |N_{wg}|) \times |\Omega| \tag{6.3}$$

$$BV = |N_n| \tag{6.4}$$

If no contingency reduction technique is used, then $|N_s^{\omega}| = |N_l| + 1$ to model outage of each line. If the VCL algorithm is used for contingency reduction, then $|N_s^{\omega}| = |ICL^{\omega}| + 1$.

6.2.2.2 Design of decomposition algorithms

PH and BD are not black-box software packages with input and output vectors. These algorithms are designed based on specific needs and conditions. For BD, there are several different designs such as standard BD [11], multi-cuts BD [15], and nested BD [41], and each design can be configured differently. For PH, either the standard form [101] or the bundled form [116] might be used. Similar to BD, there are several internal settings for PH that can affect the performance of this algorithm.

6.2.2.3 Existing hardware infrastructure

The machine that is used to solve the TEP problem has an undeniable impact on the choice of a decomposition algorithm and the size of each bundle (κ). Machines with high computing power are usually capable of solving larger problems that make it possible to choose bundled PH with a large bundle size (κ). In the case of using multiple machines (or virtual machines for Cloud based workstations), implemented parallel computation structure will be another key factor.

6.2.2.4 Solvers

The main feature of a solver that affects the choice of parameters for the framework is its capability to distribute computation over multiple cores of a CPU and use all computing power of the machine. GUROBI and CPLEX are examples of commercial solvers with this capability.

As discussed above, there are several factors that can affect hardware and software design of this framework. For a designed framework, running a few individual simulations can provide a relatively good understanding about the performance of each module, and help on setting parameters for the framework.

6.2.3 Linking PH and BD

Usually steps 3 and 5 are the most time consuming steps of the proposed framework in section 5.2.1 for large-scale problems. These steps can be solved by either HP, BD or both (Hybrid). The algorithm explained in Figure 5.3 is used as the main structure for solving TEP in steps 3 and 5. In the following, it is explained how this algorithm is appropriate for all combinations in Table 6.1. For PH (the second column of the table), the whole algorithm is run and the extensive form of stochastic TEP is solved in lines 3 and 11 in Figure 5.3. For BD (the third column of the table), BD is used to solve TEP in line 3, and the algorithm is terminated in line 4. For the hybrid case (the fourth column of the table), the whole algorithm in Figure 5.3 is run, and BD can be used to solve TEP in lines 3 and/or 11. If the BD is not used, the EF of TEP is solved. For $\kappa = |\Omega|$, Err in line 6 will be zero and the algorithm will be terminated in line 6.

6.2.4 PH performance improvement

Several heuristics such as finding appropriate values for $\boldsymbol{\rho}$, variable freezing, cyclic behavior detection, and terminating PH when the number of remaining unconverged variables is small can be used to improve the performance of the PH algorithm [115]. In the following we will discuss some of these heuristic methods that are used in this chapter.

6.2.4.1 Choice of $\boldsymbol{\rho}$

A good approximation for $\boldsymbol{\rho}$ is important for the PH algorithm to perform well. As shown in Figure 5.3, the value of multiplier vector ($\boldsymbol{W}^{v}_{\mathcal{B}_i}$) is updated using penalty vector $\boldsymbol{\rho}$, and an appropriate multiplier vector can affect the number of required iterations for PH convergence, and the quality of the lower bound answer [34]. In [115], different heuristic methods for calculating effective values for $\boldsymbol{\rho}$ are proposed. Our experience with those methods shows that for the TEP problem using the following equation from [115] results in a better convergence rate.

$$\rho_l = \frac{\zeta_l}{x_l^{max} - x_l^{min} + 1} \tag{6.5}$$

where ρ_l is the l^{th} element of vector $\boldsymbol{\rho}$, and

$$x_l^{max} = \max_{\mathcal{B}_i \in \mathcal{B}} x_l^{\mathcal{B}_i} \tag{6.6}$$

$$x_l^{min} = \min_{\mathcal{B}_i \in \mathcal{B}} x_l^{\mathcal{B}_i} \tag{6.7}$$

For values of ρ_l close to the unit cost of its associated variable, the PH algorithm should have a better performance both from convergence speed and quality of results. Selecting higher values for ρ_l will increase convergence rate but may negatively affect the quality of results. On the other hand, very small values for ρ_l can improve the quality of results (by decreasing optimality gap), but can significantly increase the number of iterations.

6.2.4.2 Variable Freezing

To improve the convergence of PH algorithm, the *variable freezing* technique can be used. Based on this technique, first stage decision variables with values that did not change over the past ϑ iterations are frozen for future iterations. For example, for a case with 5 bundles and $\vartheta = 4$, the value of the decision variable x_l is frozen if for all 5 bundles during all 4 successive iterations $\upsilon+1, \upsilon+2, \upsilon+3, \upsilon+\vartheta = \upsilon+4$, its value did not change ($x_l^{\upsilon+1,1} = \cdots = x_l^{\upsilon+4,5}$).

The impact of freezing variables can be investigated from two perspectives; i.e. its impact on simulation time and its impact on the selected plan.

- Impact on simulation time

  By freezing binary variables, the total number of binary variables is decreased as frozen variables have fixed values and no decision about them will be made in subsequent iterations. It improves the performance of the algorithm by decreasing computational time for each iteration (as a TEP optimization problem with fewer binary variables will typically solve faster) and reducing the number of iterations (as a Ph problem with fewer non-anticipativity constraints will typically converge faster).

- Impact on the selected plan

  When a decision variable is frozen, the implicit assumption is that its value will not change during subsequent iterations, but this assumption may not always be valid. Therefore, the selected plan might be negatively affected when variable freezing technique is used, especially for small values of ϑ like 1 or 2. By using more conservative values for ϑ, this effect can be mitigated.

The selected plan will be more sensitive to a small value for ϑ when there are several relatively similar candidate lines (in terms of cost and/or electric parameters) in a geographically limited area. For a large-scale network in which candidate lines are widely spread, a smaller value for ϑ can be selected.

Using the variable freezing technique may result in situations with only a very few unfrozen decision variables. Then PH can be terminated (to decrease the number of iterations), and the TEP with remaining binary variables solved in extensive form or using the BD algorithm.

6.2.4.3 Identical Parallel Candidate Lines

We have also noticed that having two (or more) identical parallel candidate lines can result in an unnecessary non-zero values of Err on lines 6 and/or 14 in PH algorithm (Figure 5.3) when only one of those lines is selected as a part of expansion plan. We recommend to slightly modify the investment cost for otherwise identical lines to break the symmetry.

6.2.4.4 Summary

The above mentioned heuristic techniques can be used to improve convergence of PH algorithm, but it may result in a higher optimality gap in step 7. In many practical cases, it is critical to get the result in a reasonable time; therefore a faster answer with a slightly higher optimality gap is usually acceptable.

6.2.5 Optimality gap

The optimality gap is used as a measure for quantifying the quality of results in an optimization-based TEP. Based on Table 6.1, the TEP problem is solved using one of these five methods i.e. heuristic, Extensive Form (EF), PH, BD, and hybrid. For parameter settings that will result in a heuristic method, we cannot calculate the optimality gap to quantify the quality of results. For the EF method, the optimality gap of the final result will be less than or equal to the solver's setting for maximum optimality gap. For BD, achieving the optimality gap is set as the stopping criterion; therefore, for EF and BD methods, it is possible to guarantee a pre-defined optimality gap (assuming that the algorithm successfully terminates). On the other hand, for PH and hybrid methods, the optimality gap is calculated after the algorithm is terminated to quantify the quality of final results, and there is no guarantee that the final optimality gap will be less than or equal to a pre-defined threshold. As discussed in section 6.2.4, using appropriate values for $\boldsymbol{\rho}$ and setting a conservative value for ϑ can improve the optimality gap of the PH algorithm.

6.2.6 Scalability and Maintainability

Scalability is one of the main features of the proposed framework. We use Figure 6.1 to discuss different aspects of this feature. Figure 6.1(a) shows the size of the EF of a stochastic TEP problem with security constraints. In this Fig., d^{ω} represents the size of the TEP problem for scenario ω ($d^{\omega} = \{SC^{\omega}, CV^{\omega}, BV^{\omega}\}$), and s is the number of scenarios ($s = |\Omega|$).

$$SC^{\omega} = 2 \times (|N_b| + |N_l|) \times |N_s^{\omega}| + |N_{wg}| \tag{6.8}$$

$$CV^{\omega} = (2 \times |N_b| + |N_l|) \times |N_s^{\omega}| + |N_g| + |N_{wg}| \tag{6.9}$$

$$BV^{\omega} = |N_n| \tag{6.10}$$

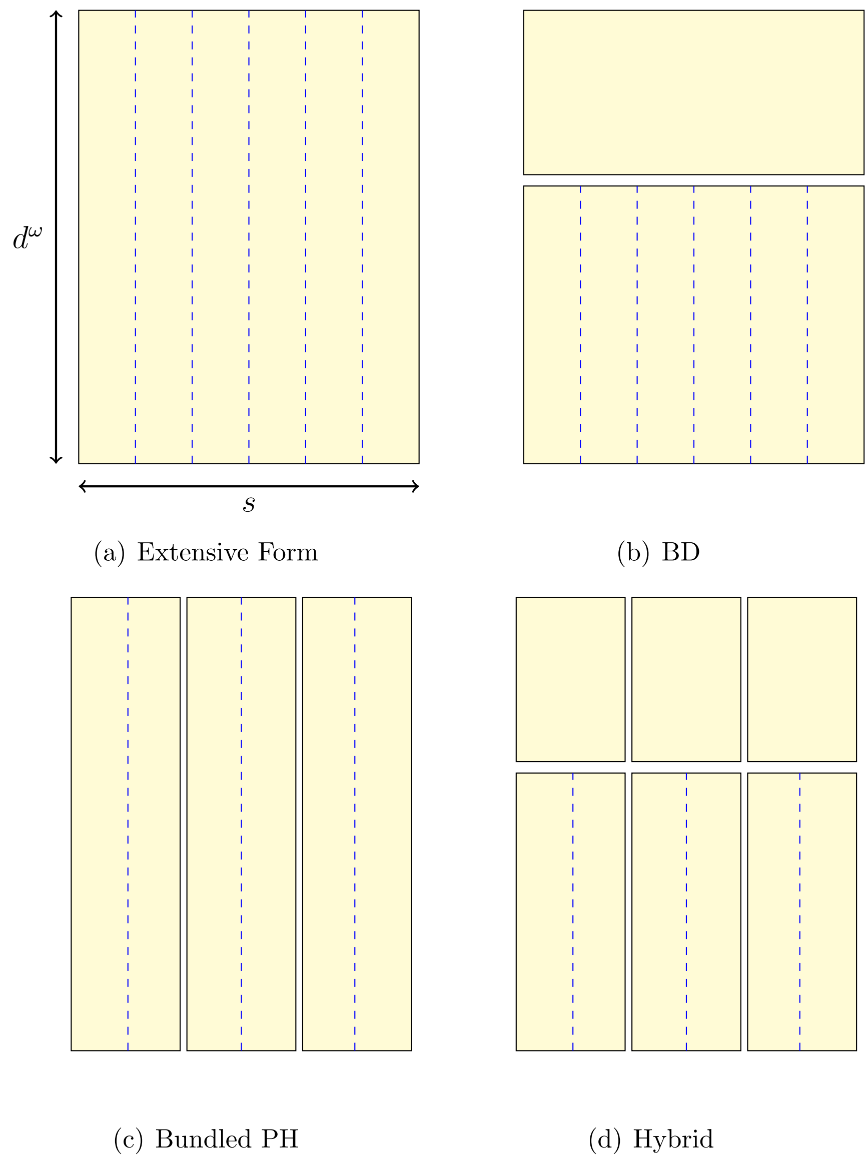
For a case system with 6000 buses, 8000 existing lines and transformers, 500 conventional power plants, 100 wind farms, 100 candidate lines and 10 scenarios, $d^{\omega} = \{228.5M, 162.8M, 100\}$ when $|N_s^{\omega}| = 8101$ and $s = 10$. Total size of the problem in Figure 6.1(a) will be $d = \{2285M, 1628M, 100\}$. This problem is practically impossible to solve in the EF. There are constraint reduction techniques [61, 64, 6] that can be used to decrease the size of this problem. Let's assume the VCL algorithm (see section 4.3.2) is used, and the size of N_s^{ω} is decreased form 8101 to 50. The size of the EF of this problem will be $d = \{14M, 10M, 100\}$. Even after a massive problem size reduction, solving the EF of the problem still remains computationally extremely expensive.

The BD algorithm (shown in Figure 6.1(b)) moves binary decision variables to the master problem, and keeps all continuous variables in the subproblem. As the subproblem is a linear program, it is expected to be solved very fast; however, for the network in this example, the size of the subproblem will be $\{14M, 10M, 0\}$ which is not easy to solve especially if it is solved in every iteration.

Figure 6.1(c) shows how bundled PH algorithm will decompose the problem. By creating bundles of two scenarios, the size of each subproblem for bundled PH will be $\{2.8M, 2.0M, 100\}$ (or $\{1.4M, 1.0M, 100\}$ for standard PH). Solving the extensive form of these subproblems might still be hard because of the large number of binary variables. In Figure 6.1(d), the hybrid methods is used to decompose the problem both vertically and horizontally. By using this method, the size of each problem that needs to be solved in EF can be decreased up to $\{1.4M, 1.0M, 0\}$, which is a significant size reduction compared to $\{14M, 10M, 100\}$ for Figure 6.1(a).

The size of this case study can be increased either by increasing the number of candidate lines or the number of scenarios. The BD feature of the hybrid method will keep us away from exponentially increasing computational time as a result of adding new binary variables, and the bundled PH feature will keep the size of each subproblem relatively unchanged even if the total number of scenarios is increased significantly (by increasing the number of bundles instead of increasing the size of each bundle). Therefore the problem remains tractable, demonstrating the scalability of the proposed framework.

Another important feature of this framework (from practicality perspective) is its maintainability. Because it is module based (BD algorithm, PH algorithm, bundling algorithm), each module can easily and (relatively)



(a) Extensive Form

(b) BD

(c) Bundled PH

(d) Hybrid

Figure 6.1: The impact of different decomposition techniques, d^{ω}: size of the problem for scenario ω, s: the number of scenarios (6 for this example)

independently upgraded as technology improves.

6.2.7 Parallelizing

With proper hardware, parallelizing decreases computational time for solving a series of independent simulations, and it improves scalability of the framework. Simulations in steps 3 and 5 in the proposed framework can be parallelized, if PH, BD (with special configurations), or hybrid is selected to reduce elapsed time for solving TEP optimization problem by starting all simulations at the same time.

6.2.7.1 PH algorithm

Based on PH algorithm for bundled scenarios shown in Figure 5.3, lines 3 and 11 are run for each bundle (or each scenario in case of standard PH) independently. Therefore, we can parallelize both `for loops` (lines 2-4 and 10-12) in this algorithm, and start all simulations in each `loop` at the same time to decrease computational time. It should be noted that lines 10-12 should be solved for each iteration of the PH algorithm, and decreasing computational time here can be rewarding from the performance improvement perspective. As shown in lines 5 and 13 in Figure 5.3, the algorithm can proceed to the next step when all parallelized simulations are completed. In the bundling process, it should be considered to develop bundles that need relatively similar computational time, so that the framework can benefit the most from parallelizing.

6.2.7.2 BD algorithm

For standard BD, in which one cut is sent to master problem in each iteration, the subproblem should be solved in extensive form. For multi-cuts BD [15] and nested BD [41], [1] and [54], it is possible to solve subproblems in parallel that will decrease computational time.

6.2.7.3 Hybrid method

As hybrid algorithm uses both PH and BD to solve a problem, it can benefit from both vertical and horizontal decomposition techniques and parallelize the problem solving with both algorithms (if applicable). For example, by using bundled PH, the problem will be vertically parallelized for each bundle $\mathcal{B}_i$. A nested BD can be used to solve each bundle, in which feasibility cuts for contingency operation states can be created in parallel.

6.3 Case Study and Numerical Results

In this section, we run numerical analysis for two case studies i.e. a 13-bus system with 100 scenarios and a reduced ERCOT system with 10 scenarios. All simulations are done with a personal computer with 2.0-GHz CPU and 32 GB of RAM. The proposed method is implemented in MATLAB R2014a [74] by using YALMIP R20150626 package [60] as a modeling software and GUROBI 5.6 [45] as a solver. To calculate the elapsed "Simulation Time," MATLAB built-in function `tic toc` is used. Steps 3 and 5 are parallelized using MATLAB built-in function `parfor`.

6.3.1 13-bus test system

This case study contains 13 buses, 33 existing lines, 16 power plants, 9 load centers, and 36 candidate lines with 100 scenarios to capture uncertainties in wind and load (from Chapter 4). This small case study with a large number of scenarios is used to demonstrate different steps of the proposed framework. The proposed method in Chapter 4 is used for solving TEP subproblems in lines 3 and 11 of the PH algorithm. To evaluate the performance of the proposed method, this test system is solved with four different methods that are explained in the following:

6.3.1.1 Case A

In case A, a standard PH (without bundling) is used to solve TEP problem. This method is used by [87] to solve TEP without contingency analysis. As stated before, MATLAB built-in function `parfor` is used to parallelize solving TEP for each scenario.

6.3.1.2 Case B

For case B, scenarios are bundled randomly using `randperm` function in MATLAB (instead of using the proposed method in section 5.2.2) to show the impact of bundling on performance of PH algorithm for TEP problem. The size of bundles is selected based on the problem size and machine's configuration ($\kappa = 20$).

6.3.1.3 Case C

This case solves the problem using the proposed framework in Chapter 5. To show the implementation of the proposed framework, all steps are explained in detail.

Step 1: for phase I, κ is set to 50 as TEP without contingency analysis is solved, so a larger number of scenarios can be bundled compared to TEP with contingency analysis. For phase II, κ is set to 20 to fairly compare the result of cases B and C. Step 2: An OPF is solved for the base case to calculate LWs for bundling. Load shedding and wind curtailment penalties are set to \$9000/MWh and \$500/MWh respectively. It will result in the weight factor of 18 for load shedding (and 1 for wind curtailment), and LW for each scenario is calculated as the weighed sum of normalized wind and load curtailment in that scenario. Based on (5.2), $\mathcal{C}_s = 2$ and scenarios are clustered with the objective of minimizing the distance between LW^{ω} values in each cluster. In the last step, members of each cluster are distributed between bundling groups to minimize the distance between aggregated LW values ($LW_{\mathcal{B}_i}$s). Step 3: bundled PH is used to solve TEP without contingency analysis in this step. The final target is to solve TEP with contingency analysis, and results of this step are used as inputs for step 4 (to calculate a bundling attribute); therefore this step does not need to be solved until optimality. If TEP without contingency analysis is the final target, this step should be solved iteratively until the stopping criteria is met.

Phase II, step 4: the VCL algorithm (developed in Chapter 2) is used to find important lines for contingency analysis (ICL^{ω}s) using results from step 3. Scenarios are classified into 4 classes based on the size of ICL^{ω}s ($|ICL^{\omega}|$). Then, scenarios in each class are clustered based on similarity/dissimilarity of their ICL lists. It will result in clusters having members with relatively similar ICLs. In the last step of bundling, scenarios in each cluster are distributed between target bundles to create groups with relatively the same number of ICLs. This criterion tries to balance computational burden between groups. The size of ICLs in each group affects the number of operation states and consequently computational time. In step 5, bundled PH is solved iteratively until stopping criteria is met.

Phase III, Step 6: A lower bound is calculated (based on the proposed method by [34]) to quantify the quality of the result from step 5. In step 7, optimality gap is calculated based on upper and lower bounds from steps 5 and 6 respectively.

6.3.1.4 Case D

For this case, the proposed method in Chapter 4 is used to solve the stochastic TEP problem with contingency analysis.

6.3.1.5 PH algorithm settings

The values of $\boldsymbol{\rho}$ are calculated based on (6.5). Freezing variables is one of the techniques that is used to improve convergence of PH algorithm. Variables that do not change over the most recent 4 iterations will be frozen at their values ($\vartheta = 4$). Moreover, if the number of remaining binary variables is less than or equal to 3, the PH algorithm is terminated, and the extensive form of the problem is solved for remaining decision variables. These settings are applied to three cases A-C.

6.3.1.6 Model performance discussion

The simulation result for these four cases is summarized in Table 6.2. Standard PH in case A needs more than 2 hours to solve this problem and the final result is 29.5%-suboptimal. It shows that standard PH will not have a good performance when the number of scenarios is large. For Case B, bundling reduced computational time by 50% and optimality gap is dropped to 1.65%. For case D, the TEP optimization problem is solved in 25 minutes with 2.7% optimality gap. The proposed method in case C reduced computational time to 15 minutes, and significantly improved the quality of results by decreasing optimality gap to 0.24%. Figure 6.2 shows how computational time (left axis-solid blue line) and optimality gap (right axis-dashed orange line) are changed from case A–D. Computational time is normalized based on total time for case C. The proposed framework solves this problem more than 8 times faster than standard PH and 5 times faster than randomly bundled PH. It also finds results with higher quality (0.24% compared to 1.65% and 29.4% for randomly bundled PH and standard PH respectively). From computational time perspective, cases C and D are relatively similar, but the quantified quality of results is significantly different, and case C provides a better optimality gap

Table 6.2: Summary of results for 13-Bus system

	Case A	Case B	Case C	Case D
No. of added lines	21	17	16	16
Objective Function ($b)	5.58	4.94	4.89	4.89
Simulation Time (hrs)	2.05	1.28	0.25	0.42
Optimality Gap	29.5%	1.65%	0.24%	2.7%

in somewhat less time.

To investigate the impact of parallelizing and variable freezing on computational time, we compared the performance of the framework under the following three alternatives:

- *Alter. 1*: With variable freezing and without parallelizing
- *Alter. 2*: Without variable freezing and with parallelizing
- *Alter. 3*: With variable freezing and with parallelizing

Table 6.3 summarizes the impact of these two factors on optimality gap and computational time for cases A-C under these three alternatives.

The result from the second row shows that variable freezing may negatively affect the quality of results and increases the optimality gap (*Alter. 2*, in which variable freezing is ignored, has the lowest optimality gap). As expected, parallelizing will not affect the quality of results (similar optimality gaps for *Alter. 1* and *Alter. 3*) The third row in Table 6.3 shows the computational time for three alternatives. For *Alter. 1*, standard PH (Case A)

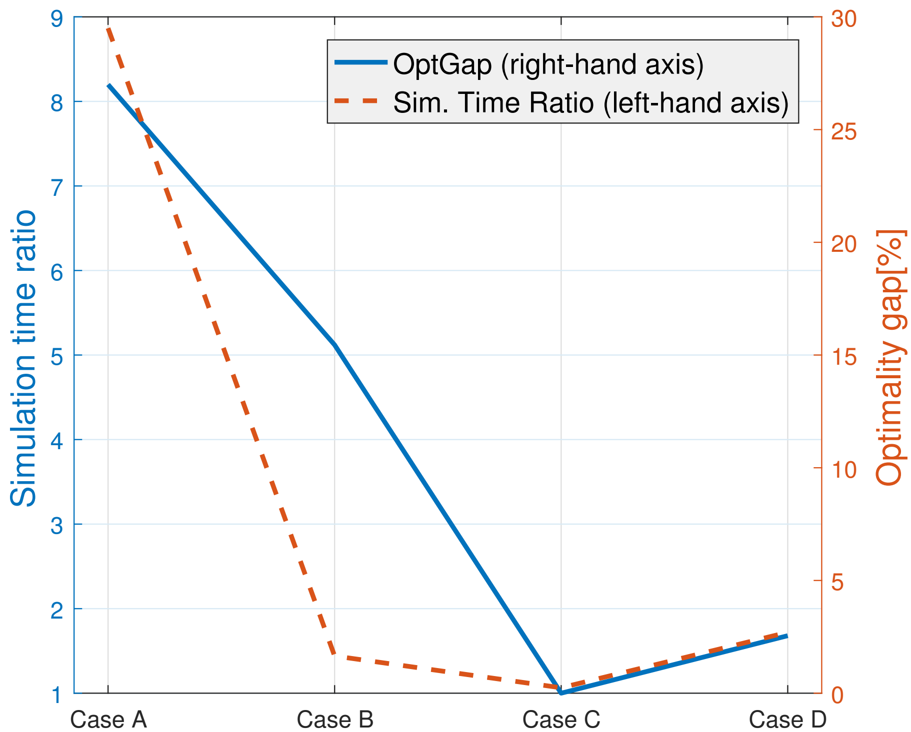


Figure 6.2: Optimality gap and the ratio of simulation time

Table 6.3: Impact of parallelizing and variable freezing on performance

		Alter. 1	Alter. 2	Alter. 3
Optimality Gap	Case A	29.5%	0.85%	29.5%
	Case B	1.65%	0.13%	1.65%
	Case C	0.24%	0.12%	0.24%
Simulation Time (hrs)	Case A	93.92	185.23	2.05
	Case B	7.38	132.97	1.28
	Case C	7.16	82.7	0.25

is affected the most (compared to cases B and C) when parallelizing is not used because each iteration includes running TEP for all individual scenarios (simulation time increased from 2.05 to 92.38 hours). For bundled PH, both cases B and C could solve the problem in relatively the same time showing that when simulations are run sequentially (instead of in parallel), the impact of balancing computational burden between bundles (that will result in an earlier termination for a parallelized `for loop`) will be less effective. Variable freezing has a significant impact on computational time as it will decrease both the number of iterations and computational time for each iteration. Comparing the computational time and optimality gap for *Alter. 2* and *Alter. 3* shows the trade-off between quality of results and computational time. For example, for case C, the optimality gap is slightly increased from 0.12% to 0.24%; however the computational time is decreased from 82.7 hours to 0.25 hours that demonstrates the effectiveness of the proposed framework.

6.3.2 ERCOT Case Study

A reduced ERCOT network is developed with 3179 buses, 474 generation units, 3598 load centers, 123 wind farms and 4458 branches. All non-radial 138kV and 345kV lines in the ERCOT network are explicitly modeled. Generators and loads that were connected to lower voltage levels or radial network are moved to nearby modeled buses. Ten different scenarios are developed to model load and wind uncertainties (using historical data) with 46 new lines as candidates for transmission expansion. Similar to the 13-bus system, four cases A–D are simulated to compare the results. For phase I in case C, $\kappa = 5$ and for case B and phase II in case C, $\kappa = 2$. The proposed method in Chapter 4 is used to solve TEP in lines 3 and 11 of the bundled PH algorithm (Figure 5.3). The parameter ϑ is set to 3. All other parameters are set the same as the 13-bus system. Numerical result is given in Table 6.4. As the number of scenarios is not large for this system, standard PH (case A) has a reasonable performance; however, the elapsed time of over a week may not be acceptable. For case B (randomly bundled scenarios), simulation is terminated manually after 14.9 days and a lower bound is calculated. The fourth column (case C) demonstrates the impact of the proposed framework on improving quality of results (decreasing optimality gap from 6.24% to 0.97%) and reducing computational time (by more than 5.3 times) for solving this large-scale problem. We could not get a feasible solution for case D after 15 days, demonstrating the need for decomposition-based methods for large-scale problems. Results for this case demonstrates that bundling by itself may not necessar-

Table 6.4: Summary of results for ERCOT system

	Case A	Case B	Case C	Case D
No. of added lines	6	9	4	–
Objective Function ($b)	8.102	8.230	8.007	–
Simulation Time (days)	9.2	14.9	2.78	15
Optimality Gap	3.1%	6.24%	0.97%	–

ily improve the performance of PH without careful consideration of choice of bundles, because as explained in section 5.3.2, each iteration for the PH algorithm is finished when TEP for all bundles are completely solved (lines 5 and 13 in Figure 5.3). Because of this, randomly grouping scenarios may result in forming TEP subproblems with significantly different sizes (based on (6.2) and (6.3)) although the size of bundles (κ) is similar. This comparison also highlights the importance of grouping step in the proposed bundling method.

6.4 Summary

In Chapters 5 and 6, a generalized decomposition framework is developed for solving large-scale TEP problems. This framework is easily scalable, and its flexible structure makes it possible to configure it for problems with different sizes. It allows decomposition of a problem both *vertically* and *horizontally*, using bundled PH and BD algorithms respectively. The designed steps in the framework makes it possible to parallelize simulation and keep TEP for large-scale systems tractable. A heuristic method is also developed to effectively bundle scenarios for PH algorithm. Its objective is to maximize similarity *between* bundles to improve the performance of the PH algorithm by speeding convergence of non-anticipativity constraints. Using this bundling heuristic decreased computational time by a factor of more than 8 and improved quality of results by reducing optimality gap from 29.5% to 0.24% for a 13-bus system with 100 scenarios. For a reduced ERCOT case study with 3179 buses and 10 scenarios, it provided a high quality result (0.97% optimality gap) in a reasonable time (2.8 days). The proposed framework makes solving TEP optimization problem for real-size networks tractable. This framework can be used by ISOs and transmission system owners for TEP.

Chapter 7

Conclusion

7.1 Conclusion

On the one hand, the electric transmission network is a critical infrastructure, and on the other hand it functions as an open access infrastructure for market participants. Investment costs for transmission expansion are eventually paid by rate payers; therefore, it is important to develop the network with the least cost or greatest welfare improvement, while keeping high reliability. Using optimization-based approaches for modeling transmission expansion planning (TEP) can guarantee optimality of results (or quantify the quality of results) while maximizing social welfare and satisfying security criterion. A drawback of optimization-based approaches for TEP optimization problem is their intractability that make them impractical for real-size networks. It motivates this research to investigate computational challenges related to optimization-based TEP formulations for large-scale problems.

First, we investigate the impact of contingency analysis on transmission expansion planning, and preliminary results show that ignoring contingency analysis in TEP may cause load shedding and huge extra operation costs. However, adding security constrains to the TEP optimization problem makes even medium size systems intractable in the absence of computational improvements.

We have developed the Variable Contingency List (VCL) algorithm to decrease the size of the problem by selecting a subset of lines for contingency analysis. This algorithm uses modified line outage distribution factors to approximate post contingency flow on lines. The lines for which outage does not cause any overload in the network can simply be removed from our contingency analysis list. Two options are designed for solving the TEP with security constraints i.e. option (i) that provides an upper bound (with known optimality gap) for TEP with less computational effort and option (ii) that provides the optimal answer. Depending on the whole planning process either option (i) or (ii) may be selected by the planning team.

Then, we moved toward stochastic TEP to explicitly integrate uncertainties in transmission planning decision making process, and proposed a heuristic method to decrease the size of candidate lines list (CLL) for high-level transmission expansion planning. The proposed method uses Line Closure Distribution Factors (LCDFs) to investigate the impact of adding each new candidate line. In this phase, the impact of contingency analysis on stochastic TEP is ignored.

Next, we have added security constraints (contingency analysis) into stochastic TEP. To be able to solve this large-scale optimization problem, we proposed an iterative framework that uses a three level filter to gradually add security constraints into the TEP optimization formulation. The three-level filtering algorithm uses developed important scenario identification index ($ISII$) and similar scenario elimination (SSE) technique to decrease the number of security constraints in stochastic TEP in a systematic and tractable way. This filter decreases structural constraints in our TEP mixed-integer programming formulation in each iteration (up to 99.8% in the first iteration), and the iterative framework adds reliability constraints into the optimization problem gradually in order to decrease the total simulation time. A lower bound is calculated to quantify the quality of results.

Finally, we propose a scalable and configurable decomposition framework for solving large-scale transmission capacity expansion planning with security constraints under uncertainties. This framework is capable of using both progressive hedging (PH) and Benders decomposition (BD) algorithms to decompose and parallelize a large-scale problem both vertically and horizontally. A scenario bundling method is also developed to create bundles through three steps i.e. classification, clustering, and grouping with the objective of maximizing similarity between bundles. This bundling method can improve both quality of results (decreasing optimality gap) and performance (reducing computational time) of the proposed framework. To verify capabilities of the proposed method, it is applied to a reduced ERCOT system with 3179 buses, 4458 branches and 10 scenarios. The numerical result for this case study shows that the proposed framework can make solving large-scale problems tractable, and provides high quality results (with less than 1% optimality gap) in a reasonable time (around 2.8 days) (see [71, 72]).

7.2 Future Work

The work presented in this dissertation can be extended in multiple directions. We have modeled uncertainties as a set of finite scenarios with known probabilities associated with them. Although for the majority of shorter-term "micro" uncertainties it is possible to use statistical methods to approximate probabilities related to each scenario, for macro uncertainties such as technology evolution, costs related to them, future generation mix/capacity, etc. usually expert knowledge is used to assign those probabilities. But in this case, there is typically a range of probabilities rather than a single probability that everybody agrees with. To handle this drawback, it is possible to use robust stochastic optimization formulation in which probabilities assigned to each scenario can be uncertain and change in a range. It finds the probabilities for each scenario in a way that the expected operation cost represents the worst case scenario. We are working on this subject and we have formulated the problem and developed some initial models.

Recently, authors in [36, 76, 105] used adaptive robust optimization [13] to model uncertainties for TEP to find an expansion plan that satisfies the worst case scenario. The work in this dissertation can be directed toward robust optimization to integrate a range of uncertainties that need to be considered for their worst case scenarios.

We believe the proposed decomposition framework and scenario bundling in Chapter 5 can be studied further in the future. We used expert knowledge to define and model attributes for the bundling process; however, it is possible to develop theoretical models to find appropriate attributes for bundling purpose. This is a problem we are already working on as one immediate direction for extending this thesis. The results of this extension can then be used to solve other problems with the same structure.

Another important direction for extension of this work is to move toward multi-stage TEP. It is in particular very important as TEP studies are done for near-term and long-term separately now that can affect both planning studies. In practice, the most critical decisions are those relating to nearest term construction, and long-term decisions will be revised in subsequent planning studies. Consequently, maintaining feasibility in the face of long-term uncertainties is the most critical characteristic of the solutions for far future decisions. By modeling multi-stage TEP, it will be possible to integrate near- and long-term TEP studies, in which we can investigate the impact of long-term uncertainties on near-term TEP results. It also provides more flexibility as we can modify plans that are made for the second or third stages later when more information is revealed.

Finally, we would like to mention that, in cooperation with LCRA, as a large transmission owner company, we have implemented the proposed method in this dissertation on an actual TEP project. The next step toward solving real-size networks will be collaborating with ERCOT to add more details to the model (based on an ISO's requirements) and evaluate the results and performance of the method.

Bibliography

[1] Tohid Akbari, Ashkan Rahimikian, and Ahad Kazemi. A multi-stage stochastic transmission expansion planning method. *Energy Conversion and Management*, 52(89):2844 – 2853, 2011.

[2] N. Alguacil, A.L. Motto, and A.J. Conejo. Transmission expansion planning: a mixed-integer LP approach. *IEEE Transactions on Power Systems*, 18(3):1070–1077, August 2003.

[3] I. Alhamrouni, A. B. Khairuddin, A. K. Ferdavani, M. Salem, and A. Alnajjar. Differential evolution algorithm for multistage transmission expansion planning based on ac load flow model. In *3rd IET International Conference on Clean Energy and Technology (CEAT) 2014*, pages 1–4, Nov 2014.

[4] Omid Alizadeh-Mousavi and Marija Zima-Bokarjova. Efficient benders cuts for transmission expansion planning. *Electric Power Systems Research*, 131:275 – 284, 2016.

[5] J. Alvarez Lopez, K. Ponnambalam, and V.H. Quintana. Generation and transmission expansion under risk using stochastic programming. *Power Systems, IEEE Transactions on*, 22(3):1369–1378, Aug 2007.

[6] A.J. Ardakani and F. Bouffard. Identification of umbrella constraints in dc-based security-constrained optimal power flow. *Power Systems, IEEE Transactions on*, 28(4):3924–3934, Nov 2013.

[7] Zeng B and L. Zhao. Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41:457–461, 2013.

[8] L. Bahiense, G.C. Oliveira, M. Pereira, and S. Granville. A mixed integer disjunctive model for transmission network expansion. *IEEE Transactions on Power Systems*, 16(3):560–565, August 2001.

[9] R. Baldick, K. Dixit, and T.J. Overbye. Empirical analysis of the variation of distribution factors with loading. In *Power Engineering Society General Meeting, 2005. IEEE*, pages 221–229 Vol. 1, June 2005.

[10] Ross Baldick. *Applied Optimization: Formulation and Algorithms for Engineering Systems*. Cambridge University Press., 2006.

[11] J. F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 3:238–252, 1962.

[12] R. Bent, C. Coffrin, R. R. E. Gumucio, and P. Van Hentenryck. Transmission network expansion planning: Bridging the gap between ac heuristics and dc approximations. In *2014 Power Systems Computation Conference*, pages 1–8, Aug 2014.

[13] Dimitris Bertsimas, David B. Brown, and Constantine Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[14] S. Binato, M. V F Pereira, and S. Granville. A new Benders decomposition approach to solve power transmission network design problems. *IEEE Transactions on Power Systems*, 16(2):235–240, May 2001.

[15] John R. Birge and Francois V. Louveaux. A multicut algorithm for two-stage stochastic linear programs. *European Journal of Operational Research*, 34:384–392, 1988.

[16] M.O. Buygi, G. Balzer, H.M. Shanechi, and M. Shahidehpour. Market-based transmission expansion planning. *Power Systems, IEEE Transactions on*, 19(4):2060–2067, Nov 2004.

[17] M. Carrion, J.M. Arroyo, and N. Alguacil. Vulnerability-constrained transmission expansion planning: A stochastic programming approach. *Power Systems, IEEE Transactions on*, 22(4):1436–1445, Nov 2007.

[18] Enrique B. Cedeo and Sant Arora. Performance comparison of transmission network expansion planning under deterministic and uncertain conditions. *International Journal of Electrical Power and Energy Systems*, 33(7):1288 – 1295, 2011.

[19] Guo Chen, ZhaoYang Dong, and D.J. Hill. Transmission network expansion planning with wind energy integration: A stochastic programming

model. In *Power and Energy Society General Meeting, 2012 IEEE*, pages 1–10, July 2012.

[20] Jens Clausen. *Branch and Bound Algorithms-Principles and Examples.* University of Copenhagen., March 1999. Technical report.

[21] C. Coffrin, P. Van Hentenryck, and R. Bent. Approximating line losses and apparent power in ac power flow linearizations. In *Power and Energy Society General Meeting, 2012 IEEE*, pages 1–8, July 2012.

[22] Carleton Coffrin and Pascal Van Hentenryck. A linear-programming approximation of ac power flows. *INFORMS Journal on Computing*, 26(4):718–734, 2014.

[23] A.J. Conejo, E. Castillo, R. Minguez, and R. Garcia-Bertrand. *Decomposition Techniques in Mathematical Programming: Engineering and Science Applications*. Springer Inc., 2006.

[24] Teodor Gabriel Crainic, Mike Hewitt, and Walter Rei. Scenario grouping in a progressive hedging-based meta-heuristic for stochastic network design. *Computers & Operations Research*, 43:90 – 99, 2014.

[25] D. A. Douglass and A. A. Edris. Real-time monitoring and dynamic thermal rating of power transmission circuits. *IEEE Transactions on Power Delivery*, 11(3):1407–1418, Jul 1996.

[26] Electric Reliability Council of Texas. Attachment A: resource adequacy action, 2013.

[27] Electric Reliability Council of Texas. ERCOT planning guide, section 4: Transmission planning criteria, 2014.

[28] J. S. Engelhardt and S. P. Basu. Design, installation, and field experience with an overhead transmission dynamic line rating system. In *Proceedings of 1996 Transmission and Distribution Conference and Exposition*, pages 366–370, Sep 1996.

[29] ERCOT System Planning. 2014 Long-Term System Assessment for the ERCOT Region, 2014.

[30] ERCOT System Planning. 2016 Long-Term System Assessment for the ERCOT Region, 2016.

[31] Yu. Ermoliev and R. J-B Wets. *Numerical Techniques for Stochastic Optimization*. Springer series in computational mathematics, 1988.

[32] ESB NationalGrid. Transmission planning criteria, 1998.

[33] Laureano F. Escudero, Mara Araceli Garn, Mara Merino, and Gloria Prez. An algorithmic framework for solving large-scale multistage stochastic mixed 01 problems with nonsymmetric scenario trees. *Computers & Operations Research*, 39(5):1133 – 1144, 2012.

[34] Dinakar Gade, Gabriel Hackebeil, Sarah M. Ryan, Jean-Paul Watson, Roger J.-B. Wets, and David L. Woodruff. Obtaining lower bounds from the progressive hedging algorithm for stochastic mixed-integer programs. *Mathematical Programming*, 157(1):47–67, 2016.

[35] L.P. Garces, A.J. Conejo, R. Garcia-Bertrand, and R. Romero. A bilevel approach to transmission expansion planning within a market environment. *Power Systems, IEEE Transactions on*, 24(3):1513–1522, Aug 2009.

[36] R. Garcia-Bertrand and R. Minguez. Dynamic robust transmission network expansion planning. *IEEE Transactions on Power Systems*, PP(99):1–1, 2016.

[37] L.L. Garver. Transmission network estimation using linear programming. *IEEE Transactions on Power Apparatus and Systems*, PAS-89(7):1688–1697, September 1970.

[38] R. B. Gilbert and M. Habibi. Assessing the value of information to design site investigation and construction quality assurance programs. In K.K. Phoon and J. Ching, editors, *Risk and Reliability in Geotechnical Engineering*, chapter 13, pages 491–532. CRC Press, Taylor and Francis Group, Boca Raton FL, 2015.

[39] R. B. Gilbert, M. Habibi, and N. Min. Achieving reliability in the face of extreme uncertainty. In *Fifth Asian-Pacific Symposium on Structural Reliability and its Applications (5APSSRA)*, pages 39–52, May 2012.

[40] R. B. Gilbert, M. Habibi, and F. Nadim. Accounting for unknown unknowns in managing multi-hazard risks. In P. Gardoni and J.M. LaFave, editors, *Multi-hazard Approaches to Infrastructure Engineering*, chapter 18, pages 386–412. Springer Int., Boca Raton FL, 2016.

[41] C. Roger Glassey. Nested decomposition and multi-stage linear programs. *Management Science*, 20(3):282–292, 1973.

[42] P. V. Gomes and J. T. Saraiva. Hybrid discrete evolutionary pso for ac dynamic transmission expansion planning. In *2016 IEEE International Energy Conference (ENERGYCON)*, pages 1–6, April 2016.

[43] S. Granville, M. Pereira, G. Dantzig, B. Ivi-Itzhak, M. Avriel, A. Monticelli, and L. Pinto. *Mathematical Decomposition Techniques for Expansion Planning Vol 2: Analysis of the Linearised Power Flow Model Using the Bender Decomposition Technique.* EPRI Technical Report EL-5299, 1988.

[44] Jiachun Guo, Yong Fu, Zuyi Li, and M. Shahidehpour. Direct calculation of line outage distribution factors. *IEEE Transactions on Power Systems*, 24(3):1633–1634, August 2009.

[45] Gurobi Optimization, Inc. Gurobi optimizer reference manual, 2014.

[46] M. Habibi, R. B. Gilbert, L. W. Lake, and P. McIntosh. Assessing value of test wells in developing an unconventional play decision entropy theory. In *SPE Hydrocarbon Economics and Evaluation Symposium*, pages 1–12, May 2014.

[47] Jiawei Han, Micheline Kamber, and Jian Pei. *Data Mining: Concepts and Techniques, Third Edition.* Elsevier Inc., 2011.

[48] K.W. Hedman, R.P. O'Neill, E.B. Fisher, and S.S. Oren. Optimal transmission switching with sensitivity analysis and extensions. *IEEE Transactions on Power Systems*, 23(3):1469–1479, August 2008.

[49] R. Hemmati, R.-A Hooshmand, and A Khodabakhshian. Comprehensive review of generation and transmission expansion planning. *Generation, Transmission Distribution, IET*, 7(9):955–964, Sept 2013.

[50] B. Hua and R. Baldick. A convex primal formulation for convex hull pricing. *IEEE Transactions on Power Systems*, 2016.

[51] R. Jiang, J. Wang, M. Zhang, and Y. Guan. Two-stage minimax regret robust unit commitment. *IEEE Transactions on Power Systems*, 28(3):2271–2282, Aug 2013.

[52] S. Jin and S. M. Ryan. A tri-level model of centralized transmission and decentralized generation expansion planning for an electricity market:part i. *IEEE Transactions on Power Systems*, 29(1):132–141, Jan 2014.

[53] P. Kall and S.W. Woodruff. *Stochastic Programming.* Wiley, Chichester, 1994.

[54] A. Khodaei, M. Shahidehpour, and S. Kamalinia. Transmission switching in expansion planning. *IEEE Transactions on Power Systems*, 25(3):1722–1733, August 2010.

[55] I. Konstantelos and G. Strbac. Valuation of flexible transmission investment options under uncertainty. *Power Systems, IEEE Transactions on*, 30(2):1047–1055, March 2015.

[56] A. H. Land and A. G. Doig. An Automatic Method for Solving Discrete Programming Problems. *Econometrica*, 28:497–520, 1960.

[57] G. Latorre, R.D. Cruz, J.M. Areiza, and A Villegas. Classification of publications and models on transmission expansion planning. *Power Systems, IEEE Transactions on*, 18(2):938–946, May 2003.

[58] G. Latorre-Bayona and I. J. Perez-Arriaga. Chopin, a heuristic model for long term transmission expansion planning. *IEEE Transactions on Power Systems*, 9(4):1886–1894, Nov 1994.

[59] A.M. Leite da Silva, L.S. Rezende, L. A F Manso, and G.J. Anders. Transmission expansion planning: A discussion on reliability and $N-1$ security criteria. In *2010 IEEE 11th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)*, pages 244–251, June 2010.

[60] J. Lofberg. Yalmip : A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.

[61] R. Madani, J. Lavaei, and R. Baldick. Constraint screening for security analysis of power networks. *IEEE Transactions on Power Systems*, PP(99):1–1, 2016.

[62] M. Majidi Q., S. Afsharnia, M. S. Ghazizadeh, and A. Pazuki. A new method for optimal location of facts devices in deregulated electricity market. In *Electric Power Conference, 2008. EPEC 2008. IEEE Canada*, pages 1–6, Oct 2008.

[63] M. Majidi Q, M.S. Ghazizadeh, and S. Afsharnia. A novel approach to allocate transmission embedded cost based on mw-mile method under deregulated environment. In *Electric Power Conference, 2008. EPEC 2008. IEEE Canada*, pages 1–6, Oct 2008.

[64] M. Majidi-Qadikolai and R. Baldick. Stochastic transmission capacity expansion planning with special scenario selection for integrating n-1 contingency analysis. *IEEE Transactions on Power Systems*, 31(6):4901–4912, Nov 2016.

[65] Mohammad Majidi Qadikolai and Saeed Afshania. An approach to determine the revenue share of each facts device under deregulated environment. *Journal of Applied Sciences*, pages 1677–1685, 2009.

[66] Mohammad Majidi-Qadikolai and Ross Baldick. Large-scale transmission capacity expansion planning under n-1 contingency analysis. In *INFORMS Annual Meeting, Philadelphia*, Nov. 2015.

[67] Mohammad Majidi-Qadikolai and Ross Baldick. Reducing the candidate line list for practical integration of switching into power system operation. In *22nd International Symposium on Mathematical Programming, (ISMP 2015), Pittsburgh, USA*, 2015.

[68] Mohammad Majidi-Qadikolai and Ross Baldick. Reducing the number of candidate lines for high level transmission capacity expansion planning under uncertainties. In *North American Power Symposium (NAPS), 2015*, pages 1–6, Oct 2015.

[69] Mohammad Majidi-Qadikolai and Ross Baldick. Integration of N-1 contingency analysis with systematic transmission capacity expansion planning: Ercot case study. *IEEE Transactions on Power Systems*, 31(3):2234–2245, May 2016.

[70] Mohammad Majidi-Qadikolai and Ross Baldick. Stochastic transmission expansion planning (TEP) with N-1 contingency analysis for large-scale network. In *INFORMS International Conference, Hawaii*, June 2016.

[71] Mohammad Majidi-Qadikolai and Ross Baldick. A generalized decomposition framework for large-scale transmission expansion planning. *IEEE Transactions on Power Systems*, PP(99):1–1, 2017.

[72] Mohammad Majidi-Qadikolai and Ross Baldick. Stochastic transmission expansion planning: Computational challenges. In *2017 INFORMS Annual Metting, Houston, 2017*, 2017.

[73] Market & Infrastructure Development, CAISO. California ISO 2016-2017 Regional Transmission Planning Process, 2016.

[74] MATLAB. *version 8.3.0.532 (R2014a)*. The MathWorks Inc., Natick, Massachusetts, 2014.

[75] Midcontinent ISO. MTEP16 MISO Transmission Expansion Plan, 2016.

[76] R. Minguez and R. Garcia-Bertrand. Robust transmission network expansion planning in energy systems: Improving computational performance. *European Journal of Operational Research*, 248(1):21 – 32, 2016.

[77] M Sadegh Modarresi, Tong Huang, Hao Ming, and Le Xie. Robust phase detection in distribution systems. In *Power and Energy Conference (TPEC), IEEE Texas*, pages 1–5. IEEE, 2017.

[78] Mohammad Sadegh Modarresi and Mahmud Fotuhi-Firuzabad. Effective design of the reward-penalty schemes to reach the highest persuasiveness in selecting investment strategies by electric distribution utilities. 2012.

[79] Mohammad Sadegh Modarresi, Mahmud Fotuhi-Firuzabad, and Hosein Mohammadnezhad-Shourkaei. Impacts of service quality regulation on distribution automation investment. In *Energy Conference and Exhibition (EnergyCon), 2010 IEEE International*, pages 373–378. IEEE, 2010.

[80] Mohammad Sadegh Modarresi and Le Xie. An operating reserve risk map for quantifiable reliability performances in renewable power sys-

tems. In *PES General Meeting— Conference & Exposition, 2014 IEEE*, pages 1–5. IEEE, 2014.

[81] A. Monticelli, M. V F Pereira, and S. Granville. Security-constrained optimal power flow with post-contingency corrective rescheduling. *IEEE Transactions on Power Systems*, 2(1):175–180, February 1987.

[82] A. Monticelli, A. Santos, M. V. F. Pereira, S. H. Cunha, B. J. Parker, and J. C. G. Praca. Interactive transmission network planning using a least-effort criterion. *IEEE Power Engineering Review*, PER-2(10):46–47, Oct 1982.

[83] R. Moreno, D. Pudjianto, and G. Strbac. Transmission network investment with probabilistic security and corrective control. *IEEE Transactions on Power Systems*, 28(4):3935–3944, Nov 2013.

[84] F. D. Munoz, B. F. Hobbs, J. L. Ho, and S. Kasina. An engineering-economic approach to transmission planning under market and regulatory uncertainties: WECC case study. *IEEE Transactions on Power Systems*, 29(1):307–317, January 2014.

[85] F. D. Munoz, Jean-Paul Watson, and B. F. Hobbs. New bounding and decomposition approaches for MILP investment problems: Multi-area transmission and generation planning under policy constraints. *Sandia National Laboratories (SNL-NM)*, (SAND2014-4398J), 2014.

[86] Francisco D. Munoz, EnzoE. Sauma, and BenjaminF. Hobbs. Approximations in power transmission planning: implications for the cost and performance of renewable portfolio standards. *Journal of Regulatory Economics*, 43(3):305–338, 2013.

[87] Francisco D. Munoz and Jean-Paul Watson. A scalable solution framework for stochastic transmission and generation planning problems. *Computational Management Science*, 12(4):491–518, 2015.

[88] Francisco D. Munoz, Jean-Paul Watson, and Benjamin F. Hobbs. Optimizing your options: Extracting the full economic value of transmission when planning under uncertainty. *The Electricity Journal*, 28(5):26 – 38, 2015.

[89] NERC. Transmission system adequacy and security, 2005.

[90] NERC. Definition of adequate level of reliability, 2007.

[91] NERC. Transmission system planning performance requirements, 2015. Standard TPL-001-4.

[92] Richard P ONeill, Eric A Krall, Kory W Hedman, and Shmuel S Oren. A model and approach for optimal power systems planning and investment. *Math. Program*, 2011.

[93] T.J. Overbye, Xu Cheng, and Yan Sun. A comparison of the AC and DC power flow models for lmp calculations. In *Proceedings of the 37th*

Annual Hawaii International Conference on System Sciences, pages 9 pp.–, Jan 2004.

[94] Heejung Park and R. Baldick. Transmission planning under uncertainties of wind and load: Sequential approximation approach. *IEEE Transactions on Power Systems*, 28(3):2395–2402, Aug 2013.

[95] M. V. F. Pereira and L. M. V. G. Pinto. Application of sensitivity analysis of load supplying capability to interactive transmission expansion planning. *IEEE Transactions on Power Apparatus and Systems*, PAS-104(2):381–389, Feb 1985.

[96] Johannes Pfeifenberger and Judy Chang. *Well-Planned Electric transmission Saves Customer Costs: Improved transmission planning is key to the transmission to a carbon-constrained future*. The Brattle Group, June 2016.

[97] Potomac Economics. 2013 active constraint list, 2014. Personal Correspondence.

[98] PowerWorld Inc. Power transfer capability limits.

[99] Rolando Pringles, Fernando Olsina, and Francisco Garcs. Real option valuation of power transmission investments by stochastic simulation. *Energy Economics*, 47:215 – 226, 2015.

[100] M. J. Rider, A. V. Garcia, and R. Romero. Power system transmission network expansion planning using ac model. *IET Generation, Transmission Distribution*, 1(5):731–742, September 2007.

[101] R.T. Rockafellar and R. J-B. Wets. Scenario and policy aggregation in optimization under uncertainty. *Mathematics and Operation Research*, pages 119–147, 1991.

[102] R. Romero and A. Monticelli. A hierarchical decomposition approach for transmission network expansion planning. *IEEE Transactions on Power Systems*, 9(1):373–380, February 1994.

[103] R. Romero and A. Monticelli. A zero-one implicit enumeration method for optimizing investments in transmission expansion planning. *IEEE Transactions on Power Systems*, 9(3):1385–1391, August 1994.

[104] A.M. Rudkevich. A nodal capacity market for co-optimization of generation and transmission expansion. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1080–1088, October 2012.

[105] C. Ruiz and A.J. Conejo. Robust transmission expansion planning. *European Journal of Operational Research*, 242(2):390 – 401, 2015.

[106] P.A. Ruiz, J.M. Foster, A. Rudkevich, and M.C. Caramanis. Tractable transmission topology control using sensitivity analysis. *IEEE Transactions on Power Systems*, 27(3):1550–1559, August 2012.

[107] P.A. Ruiz, A. Rudkevich, M.C. Caramanis, E. Goldis, E. Ntakou, and C.R. Philbrick. Reduced MIP formulation for transmission topology control. In *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1073–1079, October 2012.

[108] S. M. Ryan, R. J. B. Wets, D. L. Woodruff, C. Silva-Monroy, and J. P. Watson. Toward scalable, parallel progressive hedging for stochastic unit commitment. In *2013 IEEE Power Energy Society General Meeting*, pages 1–5, July 2013.

[109] P.W. Sauer, K.E. Reinhard, and T.J. Overbye. Extended factors for linear contingency analysis. In *Proceedings of the 34th Annual Hawaii International Conference on System Sciences*, pages 697–703, January 2001.

[110] Henry C. Thode. *Testing For Normality*. CRC Press, 2002.

[111] Dirk Van Hertem, Jody Verboomen, Konrad Purchala, Ronnie Belmans, and WL Kling. Usefulness of DC power flow for active power flow analysis with flow controlling devices. In *AC and DC Power Transmission, 2006. ACDC 2006. The 8th IEE International Conference on*, pages 58–62. IET, 2006.

[112] R. V. Villanasa. *Transmission network planning using linear and mixed linear integer programming*. PhD thesis, Ressenlaer Polythechnic Institute, 1984.

[113] R. Villasana, L.L. Garver, and S.J. Salon. Transmission network planning using linear programming. *IEEE Transactions on Power Apparatus and Systems*, PAS-104(2):349–356, February 1985.

[114] J.C. Villumsen, G. Bronmo, and A.B. Philpott. Line capacity expansion and transmission switching in power systems with large-scale wind power. *IEEE Transactions on Power Systems*, 28(2):731–739, May 2013.

[115] Jean-Paul Watson and David L. Woodruff. Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. *Computational Management Science*, 8(4):355–370, 2011.

[116] Roger J.B Wets. The aggregation principle in scenario analysis and stochastic optimization. *Algorithms and Model Formulations in Mathematical ed. S. Wallace, Springer-Verlag*, pages 91–113, 1989.

[117] Allen J. Wood and Bruce F. Wollenberg. *Power Generation, Operation, and Control*. Wiley-Interscience, 2 edition, 2006.

[118] Jun Wu and K.W. Cheung. On selection of transmission line candidates for optimal transmission switching in large power networks. In *2013 IEEE Power and Energy Society General Meeting (PES)*, pages 1–5, July 2013.

[119] H. K. Youssef and R. Hackam. New transmission planning model. *IEEE Transactions on Power Systems*, 4(1):9–18, Feb 1989.

[120] Hui Zhang, G.T. Heydt, V. Vittal, and J. Quintero. An improved network model for transmission expansion planning considering reactive power and network losses. *Power Systems, IEEE Transactions on*, 28(3):3471–3479, Aug 2013.

[121] Hui Zhang, V. Vittal, and G. Heydt. An aggregated multi-cut decomposition algorithm for two-stage transmission expansion planning problems. In *Power Energy Society General Meeting, 2015 IEEE*, pages 1–5, July 2015.

[122] Hui Zhang, V. Vittal, G.T. Heydt, and J. Quintero. A mixed-integer linear programming approach for multi-stage security-constrained transmission expansion planning. *IEEE Transactions on Power Systems*, 27(2):1125–1133, May 2012.

[123] O. Ziaee and F. Choobineh. Location-allocation of tcsc devices on a power system considering n-1 contingency. In *2016 IEEE Power Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5, Sept 2016.

[124] O. Ziaee and F. Choobineh. Stochastic location-allocation of tcsc devices on a power system with large scale wind generation. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pages 1–5, July 2016.

[125] O. Ziaee and F. F. Choobineh. Optimal location-allocation of tcsc devices on a transmission network. *IEEE Transactions on Power Systems*,

32(1):94–102, Jan 2017.

[126] O. Ziaee, O. Alizadeh Mousavi, and F. Choobineh. Co-optimization of transmission expansion planning and tcsc placement considering the correlation between wind and demand scenarios. *IEEE Transactions on Power Systems*, PP(99):1–1, 2017.