ABSTRACT

KE, XINDA. Solution Methods of Large-Scale Power System Resource Scheduling and Dispatch Problems. (Under the direction of Dr. Ning Lu).

Power grid infrastructure is undergoing unprecedented changes with an increasing number of renewable generation resources and distributed energy resources being integrated into the grid from both the transmission and distribution levels. Uncertainties, variabilities, and an explosion of system variables and their operational constraints caused by the consideration of high penetration of renewable energy resources and distributed energy resources are making the power system scheduling and dispatch problems one of the hardest optimization problems to formulate and solve with satisfactory optimality, robustness, and speed. Those new challenges in problem formulation and solving process must be addressed for the economical, reliable, and safe operation of the modern power grid.

Therefore, in this thesis, I focused my effort on developing and advancing the problem formulation and solution methodologies for large-scale power system resource scheduling and dispatch problems. First, I proposed a problem reformulation for dimension reduction of the Mixed Integer Linear Programming Unit Commitment (MILP-UC) problems and a two-step hybrid solving method to improve the solving speed of UC problems. The validation of the proposed methods was performed by benchmarking their performance against a commercial unit commitment software package, PROMOD. Using the developed tool, I was able to quantify the impacts of heat waves on power system reliability and production cost at much higher time resolution for a broader range of scenarios. Second, I extended the power system resources scheduling and dispatch problem from centralized generation resources to decentralized aggregator-controllable energy resources using the
dispatch of plug-in electric vehicles (PEV) as an example. A real-time, greedy-index based dispatching policy (GIDP), which allows a PEV service aggregator to account for reward-of-service, penalty-of-service-delays, and compensation for accelerated PEV battery degradation, was proposed to dispatch and dispatch PEVs for providing ancillary service. Simulation results show that the proposed GIDP can be used by the PEV service aggregators to provide ancillary services that are equivalent to those provided by the generation resources. This will allow us to formulate the scheduled and dispatch of the distributed energy resources into the generation scheduling process.

The developed methodologies have been implemented as components of the Energy Operation Model (EOM), an open source production tool developed at the Pacific Northwest National Laboratory. My future research direction is to develop co-optimization methods for managing hydro power plants considering water management constraints, thermal-hydro coordination, as well as renewable integration needs.
Solution Methods of Large-Scale Power System Resource Scheduling and Dispatch Problems

by

Xinda Ke

A dissertation submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Electrical Engineering

Raleigh, North Carolina

2016

APPROVED BY:

_______________________________  ________________________________
Dr. Ning Lu  Dr. David Lubkeman
Committee Chair

_______________________________  ________________________________
Dr. Yunan Liu  Dr. Mo-Yuen Chow

_______________________________
Dr. Michael CW Kintner-Meyer
DEDICATION

To My Father, Mother and My Step Mother

Changmeng Ke

Chunmei Tang

Huimei Zhan
BIOGRAPHY

The author, Xinda Ke, was born in Hubei, China. He received B.S. degree in Electrical Engineering from Wuhan University of Science and Technology in 2010, and M.S. degree in Electrical Engineering from Lehigh University in 2013. He started to pursue the Ph.D. degree at the Electrical and Computer Engineering Department, North Carolina State University in 2013. His research interests include addressing climate change impact on power system planning and managing plug-in electric vehicles and energy storage devices for renewable energy integration.
ACKNOWLEDGEMENTS

I would like to express my first and most sincere gratitude to my advisor, Dr. Ning Lu, for the immeasurable amount of support throughout my Ph.D. research study. Her inexhaustible patience and astute criticism have allowed me to discover my interests in scientific research and professional career. Aside from the continuous guidance on my research, Dr. Lu provided me many detailed advices on technical writing and presentation. I am very thankful to her for accepting me as her Ph.D. student three years ago when I was in one of the hardest times in my life.

Besides my advisor, I would like to thank the rest of my graduate committee: Dr. Moyuen Chow, Dr. David Lubkeman, and Dr. Yunan Liu, for their valuable and insightful guidance on my thesis. They have been my role models to inspire and encourage me to pursue scientific research. I am also grateful to Dr. Kintner-Meyer Michael, a member of my graduate committee as well as my mentor at Pacific Northwest National Laboratory (PNNL), who provided me with the opportunity to work closely with the PNNL researchers, exposed me to many cutting edge research projects, and continuously supported me in pursuing new research directions and ideas.

I would also like to express my thanks to my dear labmates: Maziar Vanouni, Jiahong Yan, Xiangqi Zhu, Jian Lu, Weifeng Li, David Mulcahy, Gonzague Henri. It’s my great fortune to work with them during my graduate study.

At last, I am indebted to my father Changmeng Ke and my step mother Huimei Zhan, for allowing me to chase my dreams and all the sacrifices they have made for me.
TABLE OF CONTENTS

CHAPTER 1  Introduction .................................................................1

1.1  An Overview of the New Trends in Power System Technology Developments .. 1
1.2  Numerical Optimization Methods .................................................. 2
1.3  Technical Challenges ...................................................................... 3
1.4  Overview of Thesis Framework ...................................................... 5
1.5  Contributions of the Thesis ............................................................ 8
1.6  Organization of the Thesis ............................................................. 9

CHAPTER 2  State of the Art: Optimization Methods of Resource
Scheduling Problem .........................................................................11

2.1  Deterministic Optimization Methods .............................................. 11
    2.1.1  Priority List Methods ............................................................ 11
    2.1.2  The Lagrangian Relaxation Based Methods .......................... 12
    2.1.3  The Branch and Bound Method ............................................ 13
2.2  Non-Deterministic Optimization Methods ...................................... 15
    2.2.1  Introduction ........................................................................ 15
    2.2.2  The Multi-Armed Bandit Problem ....................................... 15
2.3  Solution Methods of UC Problem .................................................. 17

CHAPTER 3  A Computationally Efficient MILP Formulation for
Solving Regional Grid Scheduling Problem ......................................19

3.1  Nomenclature .............................................................................. 19
    3.1.1  Parameters .......................................................................... 19
    3.1.2  Variables ............................................................................ 20
3.2  Network Model and Simulation Environments ................................ 20
    3.2.1  Network Models ................................................................. 21
    3.2.2  Simulation Environments ..................................................... 21
3.3  MILP Reformulation of UC problem ............................................. 22
    3.3.1  Reformulation of Start-Up Cost and Shut-Down Cost Function .... 23
    3.3.2  Reformulation of Piece-Wise Production Cost Function .......... 25
LIST OF TABLES

Table 3-1. Comparison of MILP-UC problem scale before and after the reformulation. ....... 29
Table 3-2. Comparison of MILP-UC problem solution time before and after reformulation. 30
Table 5-1. Case definition........................................................................................................ 59
Table 6-1. The greedy-index based dispatching policy (GIDP). ......................................... 86
Table 6-2. An example of the customer charging request. .................................................. 89
Table 6-3. Parameters and charging requirements of the 3 PEVs. ....................................... 90
Table 6-4. Charger rated power and mix. ............................................................................... 97
Table 6-5. Scenarios of charging control. ............................................................................. 97
Table 6-6. Dispatch logic of the heuristic greedy algorithm. ................................................. 99
# LIST OF FIGURES

- Figure 1-1. EOM modeling framework ................................................................. 5
- Figure 2-1. Example of Branch and Bound tree for ILP problem 2.3 .................. 14
- Figure 3-1. Footprint of Eastern Interconnection .................................................. 21
- Figure 3-2. Log data of formulated MILP-UC in CPLEX .................................... 23
- Figure 3-3. Conceptual illustration of piecewise linear approximation of production cost curve ................................................................. 25
- Figure 3-4. Log data of reformulated MILP-UC in CPLEX ................................ 29
- Figure 4-1. Percentage of generators by the on/off status over the first week of January in 2010 for EIC system with 39 zones and 4622 generators .................. 33
- Figure 4-2. Demand curve vs. unit commitment results .................................... 34
- Figure 4-3. Production cost by category ............................................................... 35
- Figure 4-4. Sensitivity of production cost on different type of operational cost ........ 36
- Figure 4-5. Cumulative generation capacity as a function of operation related costs by category of commitment status ............................................... 37
- Figure 4-6. Flow chart of hybrid UC method ......................................................... 39
- Figure 4-7. Number of generators with commitment status predetermined in different cases ................................................................. 41
- Figure 4-8. Total production cost in different cases ............................................. 42
- Figure 4-9. EOM solution time in different cases ................................................. 42
- Figure 4-10. The EOM modeling framework ....................................................... 44
- Figure 5-1. Illustration of impacts of heat waves on power system operation ....... 47
- Figure 5-2. Monthly capacity from GT and CCGT in EIC in 2006 ......................... 50
- Figure 5-3. Temperature during a heat wave in 2000 vs. average value in Houston 51
- Figure 5-4. Capacity derating rate as a function of ambient temperature ............. 52
- Figure 5-5. Normalized heat rate as a function of ambient temperature .............. 53
- Figure 5-6. Temperature and demand in AEP (zone 1) in baseline and heat wave heat wave scenarios from 7/18/2006 to 7/24/2006 .................................................. 55
- Figure 5-7. Hourly derated capacity in EOM vs monthly capacity adjustment in PROMOD ................................................................. 56
- Figure 5-8. Derated capacity of CCGT and GT units in EIC from 7/15/2006 to 7/24/2006 . 57
- Figure 5-9. Zonal reserve margin in regular summer days and heat wave days .......... 58
- Figure 5-10. LMPs in zone 2 ALWSTTA during 7/19/2006~7/24/2006 .................. 60
Figure 5-11. LMP in EOM and PROMOD Case 1-3 during 7/12/2006~7/24/2006. .......... 61
Figure 5-12. Production costs in Case 2-4 compared with Case 1 by generator category. .... 61
Figure 6-1. Components of the EOM tool. ............................................................................. 69
Figure 6-2. Configuration of an aggregator-controlled PEV charging system for providing regulation services. ................................................................................................................. 70
Figure 6-3. Number of charging/discharging cycles versus DOD and (b) Charging and discharging cycles at DOD=50% and DOD=100% [79]. ...................................................................................... 75
Figure 6-4. Steps of calculating the equivalent charge cycles. .............................................. 76
Figure 6-5. The comparison of plan A and plan A*. .............................................................. 77
Figure 6-6. Conceptual illustration of where \( s'_i = 95\% \). .................................................... 79
Figure 6-7. Comparison of different cost parameters of different PEVs. ......................... 91
Figure 6-8. Costs of discharging, idling, and charging and the GIDP index. ...................... 92
Figure 6-9. Regulation signal over a period of 10 minute.................................................... 93
Figure 6-10. Comparison of the charging trajectories of proposed GIDP method. ............ 94
Figure 6-11. Comparison of the total profit........................................................................... 95
Figure 6-12. 4MW Regulation signal in California ISO used for comparison and analysis. . 98
Figure 6-13. Aggregator profit calculated by GIDP and HGA for the overnight charging case. ................................................................................................................................. 100
Figure 6-14. Comparison of aggregator profit from GIDP and HGA in daytime charging case................................................................................................................................. 100
Figure 6-15. Comparison of battery degradation cost from GIDP and HGA in scenario 1. 102
Figure 6-16. Comparison of battery degradation cost from GIDP and HGA in scenario 2. 102
Figure 6-17. Comparison of payment to V2G vehicles and non-V2G vehicles from GIDP and HGA in scenario 2. ........................................................................................................... 104
Figure 6-18. Profit of GIDP and HGA under different regulation service prices.............. 106
Figure 6-19. Profit of GIDP and HGA under different battery unit costs. ......................... 106
Figure 7-1. Modeling framework of EOM ............................................................................ 108
CHAPTER 1  Introduction

This chapter introduces the background of the research. New trends, state-of-the-art, challenges, and summary of the thesis are discussed in next few sections.

1.1 An Overview of the New Trends in Power System Technology Developments

The U.S. power grid is undergoing significant evolution as a result of policies encouraging the growth of renewable and distributed energy resources, emphasis on resilience due to extreme weather events, and increasing involvement of electricity customers and businesses in both managing and producing energy [1, 2]. Many new devices, such as smart meters, rooftop PV and electric vehicles, make it possible to incorporate more intermittent, renewable sources of electricity and make automatic changes that reduce transmission loss [3]. Furthermore, each of these new power devices can be distributed around the load centers to increase the reliability, which adds another degree of flexibility while also increase the complexity to the current power system. Some of the key requirements for future power grids can be summarized as follows [4]:

- Easiness of renewable integration to address global climate change.
- Electrification of transportation for reducing dependence on hydrocarbon fuels.
- Active customer participation in energy efficiency programs and grid services to enable flexible energy consumption while maintaining user privacy and comfort.
- Coordinative operation of distributed energy resources via advanced metering infrastructure and sensor networks for low cost, highly reliable, and sustainable power grid operation.

To meet those requirements and manage the increasingly high penetration of renewable resources and distributed energy resources, highly efficient energy management systems at both centralized and distributed levels are needed. At the core of those energy management systems, are algorithms that can economically schedule and dispatch resources based on their operational constraints to meet power system operational and planning needs. Therefore, I selected the topic of my Ph.D. thesis in the area of finding solution methods of large-scale power system resource scheduling and dispatch problems.

1.2 Numerical Optimization Methods

Numerical optimization plays a key role in almost every aspect of the operation and planning of modern electric power networks [5, 6]. For example, optimal power flow problems are solved to determine how to adjust the system to minimize cost while maintaining security requirements in daily operation [7]. Unit commitment (UC) problems are solved to determine the operation schedule of the generating units at every hour interval with varying loads under different constraints and environments [8, 9]. Network expansion problems are solved to determine how to expand or upgrade a power network to meet future demands [10]. In fact, the power system scheduling and dispatch problems are essentially numerical optimization problem with objective function to reduce the cost while meeting system operational constraints.
In the past few decades, many classical deterministic optimization methods have been developed and successfully applied to solve those problems, e.g., Priority List (PL) methods [11, 9], Lagrangian Relaxation (LR) methods [12, 13], Branch-and-Bound (BB) methods [8, 9, 14] for Integer Programming and Dynamic Programming (DP) methods [15, 16]. Millions of dollars can be saved annually by merely a small improvement of the solution quality obtained from those numerical optimization methods in electric power networks [17]. In recent years, a number of non-deterministic optimization methods have been developed for power grid operation and planning study, such as particle swarm optimization [18, 19], fuzzy optimization [20], robust optimization [21], etc. Compared with traditional deterministic methods, these methods have become popular because they have a theoretical advantage with respect to handling nonconvexity, uncertainty and integer variables [22].

1.3 Technical Challenges

Meeting the economy and reliability objectives in modern grids is becoming increasingly more challenging due to various reasons:

- The increasing *uncertainty* and *variability* energy caused by renewable energy sources and distributed energy supplies. Decisions must often be taken in the face of the unknown. Actions decided upon in the present will have consequences that can’t fully be determined until a later stage. The uncertainty frequently leads to very large-scale optimization models. Decision under uncertainty is further complicated in the stochastic sequential multi-stage scenario.

- The *distributed* nature of thousands and maybe millions small generation and storage devices are hard (if even possible) to schedule using traditional UC methods.
Different ownership of devices require a uniform pricing and rewarding mechanism built into the scheduling and dispatch algorithms for considering the cost of providing service. This is fundamentally different from production cost (considering mainly the fuel costs) based approaches or the most recent market-bid based mechanisms.

In the past, power grid energy management schemes focus mainly on scheduling and dispatch thermal generators using conventional UC methods. In the future, with the number of generators and controllable devices on the order of thousands and maybe millions (e.g., switchable shunt elements, PHEV batteries, rooftop PV, wind turbines, etc.), the resulting optimization problems must be computationally efficient, robust, and formulated in a flexible way to facilitate dimension reduction. Therefore, in this thesis, I focused my research on two fronts: dimension reduction of traditional UC problems and integration of scheduling distributed energy resources in UC problems. The first contribution of my research is to increase the solution speed of traditional UC problems by developing a two-step solving process such that the dimension of the original UC problem can be greatly reduced in the first step, which allow the problem be solved quickly in the second step. The second contribution of my research is to integrate the scheduling of distributed energy resources (using electrical vehicles as an example) into the UC solution process by developing a cost distribution scheme that rewards the service by performance and compensates for wear-and-tear and lost-of-comfort. The resulting algorithms are implemented as part of the Energy Operation Model developed at Pacific Northwest National Laboratory for long-term planning studies.
1.4 Overview of Thesis Framework

This research is supported by Pacific Northwest National Laboratory as part of their effort of developing an Energy Operation Model (EOM) for addressing the impact of climate change on power grid operation and planning. The modeling framework includes a Regional Earth System Model (RESM), a Building Energy Demand Model (BEND), and the EOM that consists of a network model and a unit commitment engine, as shown in Figure 1-1.

![Diagram of EOM modeling framework](image)

- **Regional Earth System Model (RESM)**: Preparing inputs for the power system planning model: temperature, stream flows, solar, wind.
- **Building Energy Demand Model (BEND)**: Preparing inputs for power system demand based on building energy use projection.

**Figure 1-1. EOM modeling framework**

The RESM is a state-of-the-art climate modeling developed at Pacific Northwest National Laboratory (PNNL) which allows air-sea interactions to be represented at regional scale. It includes weather research and forecasting (WRF) model, Community Land Model (CLM) as well as the Regional Ocean Modeling system (ROMS). In this study, the RESM is used to produce hydro generation dispatch as well as ambient temperature data to EOM and BEND in order to study the impacts of ambient temperature on power grid operation.
The BEND is a comprehensive building model developed at PNNL that simulate the thermal response of building stock at regional scale on an hourly basis. In our study, The BEND model is used to provide Temperature dependent load curves.

EOM is a C++ based open source production cost model developed at PNNL to serve as an open source production cost tool for renewable integration studies. EOM repeatedly and automatically formulates the UC problems as MILP problems and solves them by calling CPLEX (http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/) through an evaluation period that can be customized [9]. It generates the production cost, generations by plant and category, fuel usage, locational marginal price (LMP) etc.

This thesis presents the author’s work on developing and advancing the solution methodologies used in EOM for solving large-Scale power system resource scheduling and dispatch problems. The connections of each chapter in the thesis to the EOM tool development and validation case studies are shown in Figure 1-1.

Using the information and database provided from RESM, BEND and PROMOD IV, the EOM production cost models can be coupled to temperature impacts and temperature–dependent load model. The coupled system has the capability to represent the impacts of hourly temperature on load condition and available capacity and efficiency of combustion turbines, and therefore capture the potential impacts on system reliability and production cost. However, for a realistic power system like the Eastern Interconnection (EIC), the resulting UC problem is computationally intensive for the state-of-the-art implementations of UC methods and current computing capabilities. Furthermore, in order to address the hydro dispatch and uncertainties in demand curves, hundreds and thousands of UC scenarios need
to be solved quickly enough for operator planning analysis. Therefore, a MILP-UC reformulation and simplification methods has been proposed in Chapter 3 and Chapter 4 to improve the solving speed of UC problem. After the problem reformulation and simplification, a case study is presented in Chapter 5 for addressing the impacts of heat waves on power system reliability and production cost.

As shown in Chapter 5, the heat waves could lead to severe power shortage problems during the peak hours. On the other hand, the large amount of controllable distributed energy devices and associated communications systems in on distribution network of modern electric power network environment provides potential capabilities for such power shortage problem. Therefore, In Chapter 6, a real-time, greedy-index based dispatching policy (GIDP) is developed to extend the EOM capability to schedule distributed energy resources (providing ancillary services using plug-in electric vehicles (PEV)). However, different from the traditionally UC problem which is mostly solved by using centralized algorithms, the optimal PEVs dispatch problem is extremely hard to solve because of the huge amount of state variables caused by high dimensional space and uncertainty in PEVs arrival/departure times, as well as the unpredictable frequency regulation signals. The proposed GIDP simplifies the solving process by transforming the optimization problem from high-dimensional space with uncertainty to a new one-dimensional space while preserving the optimum of the original problem. By solving the transformed problem in the one-dimensional space repetitively in real-time, the obtained solutions could be inversely transformed to the original optimization problem with high-dimensional space and proved to be global optimal for both spaces. In addition, a new benefit mechanism is proposed for
PEVs based on the amount of regulation service they provide, compensation for delayed-charging, and reduction of battery lifetime because of battery degradation.

1.5 Contributions of the Thesis

In this thesis, I focused my effort on developing computationally efficient and robust solution methods for solving large-scale power system resource scheduling and dispatch problems. The contributions of my thesis are summarized as follows:

- In Chapter 3, a new MILP formulation has been proposed to solve the UC problem in large-scale power grid. The proposed formulation requires fewer binary variables and constraints than previously models, yielding a significant computational saving.

- In Chapter 4, based on pattern analysis of UC results and sensitivity of UC results to different costs, I presented a modified priority list (MPL) based MILP method that truncates the vast option space of the binary decision variables (always “on” generators and always “off” generators) to a smaller set of variables that can be efficiently solved in a MILP solver.

- In Chapter 5, I validate the performance of the developed algorithms using a case study to quantify climate change impacts. This study uses inputs from climate modeling tool to project the severity and frequency of extreme weather events such as heat waves and drought. The main contributions of this Chapter are twofold. First, I model the derating impacts on gas turbine (GT) and combine cycle gas turbine (CCGT) by adjusting the output capacity and efficiency and modify the MILP-UC formulation to incorporate the derating model. Second, I implement the modified UC
formulation and quantify the heat wave impacts on Eastern Interconnection network from system reliability and economics perspectives.

- In Chapter 6, I extend the scheduling and dispatch problem from generation resources to distributed energy resources: dispatching plug-in electric vehicles. To mitigate the shortage of generation resources, one can use the large amount of controllable distributed energy devices for helping support electric system reliability during peak load hours. I present a real-time, greedy-index based dispatching policy (GIDP) for dispatching plug-in electric vehicles (PEVs) to maximize the aggregator’s profit by providing ancillary services. The main contributions of the chapter are twofold. First, I proposed a framework for an aggregator to dispatch PEV resources, prioritize their operation, and reward them based on performance and cost when providing additional services. Second, the two contradictory considerations in the PEV dispatch problem formulations is resolved in the proposed policy: to find the global optimum of a large-scale nonconvex optimization problem and to obtain the solution in real time. The GIDP method can easily be extended to solve dispatching problems for other distributed energy resources. To the best of my knowledge, such approach has not been proposed for power system applications before.

1.6 Organization of the Thesis

The rest of this thesis is organized as follows. Chapter 2 introduces the background and current methods used for resource scheduling problem. Chapter 3 presents a reformulation of the MILP-UC model for improving the computation speed of UC problems. In Chapter 4, I proposed a two-step MPL based MILP method to further improve the computational speed of UC programs while preserving its optimality. Using the developed
production cost model, the impacts of heat waves on power grid system reserve and production cost is analyzed in Chapter 5. Chapter 6 presents a real-time, greedy-index based dispatching policy (GIDP) for dispatching plug-in electric vehicles (PEVs) to provide ancillary services. The final conclusions and some future research directions are presented in Chapter 7.
CHAPTER 2  State of the Art: Optimization Methods of Resource Scheduling Problem

This section introduces the background of the optimization methods and models to be used throughout this research.

2.1 Deterministic Optimization Methods

The deterministic solution methods for resource scheduling and dispatching problems include Priority List (PL) methods [11, 9], Lagrange Relaxation (LR) methods [12, 13], Branch-and-Bound (BB) methods [8, 9, 14] and Dynamic Programming (DP) methods [15].

2.1.1 Priority List Methods

The PL-based methods solve UC problems based on a predetermined commitment list based on production costs or operation constraints. For example, units are committed in an ascending order based on their production costs so that the most economical units are committed first [11, 9]. A simple shut-down rule could be obtained after an exhaustive enumeration of all unit combinations at each load level. The PL-based methods are simple, fast, and robust. However, they are heuristic-based algorithms and cannot guarantee the optimality of the solution. Therefore, the obtained generation schedules generally have higher operation cost than those produced by other UC solution methods. This disadvantage becomes rather common and significant, especially when operating modern large-scale power systems with many units within similar production cost ranges. Therefore, more rigorous mathematic approaches are needed to produce more economical solutions for UC problems.
2.1.2 The Lagrangian Relaxation Based Methods

The LR-based method is one of the most important optimization methods for solving resource scheduling problem. Compared with other methods, the LR-based methods have mathematical advantage because the Lagrangian multiplier provides measure of the solution quality quantitatively [12, 13]. For example, let $f(x)$ be the objective function and $G(x)$ be the set of all binding constraints.

\[
\min_{x} f(X) \\
\text{st.} \\
G(X) = 0
\]  

(2.1)

Apply the LR method to (2.1), we have

\[
L(\lambda) = \max_{\lambda} \left[ \min_{x} f(X) + \lambda^T G(X) \right]
\]  

(2.2)

where $\lambda$ is the array of Lagrange multiplier for all binding constraints. All elements in Lagrange multiplier $\lambda$ represent the marginal value the objective function will increase if there is any change on corresponding binding constraints. The solution of (2.2), which serves as the dual solution of original problem (2.1), provides an upper bound for original UC problem in (2.1). Based on the strong LP duality, there exists a multiplier $\lambda^*$ such that $L(\lambda^*) \geq f(X^*)$. If $\lambda^* \neq 0$, the LP complementary slackness will ensure that an optimal solution of $L(\lambda^*)$ must satisfy the binding constraints in (2.1) and therefore will also be optimal for $f(x)$. Unfortunately, duality theory has shown that for nonconvex problems there will be a duality gap between the cost obtained by solving the relaxed problem and the optimal cost of the original problem. Due to the non-convexity of most resource scheduling problems such as
the UC problems, the dual solution seldom satisfies both the power balance constraints and the reserve constraints.

2.1.3 The Branch and Bound Method

In a mixed-integer linear programming (MILP) problem, its integer variables present most of the solving challenges [8]. The BB method is a technique to solve a discrete variable problem by solving a sequence of simpler problems derived from the original problem. The search is organized via a branch-and-bound tree. The solution of each problem on the tree yields a lower bound on the solutions of the problems that are descendants of the fractional optimal problem in the branch-and-bound tree.

For example, for an ILP problem shows in (2.3)

\[
\begin{align*}
    z &= \max 4x_1 + 5x_2 \\
    \text{s.t} & \quad x_1 + 4x_2 \geq 5 \\
    & \quad 3x_1 + 2x_2 \geq 7 \\
    & \quad x_1 \geq 2 \\
    & \quad x_2 \geq 1 \\
    & \quad x_1, x_2 \in \mathbb{Z}_+
\end{align*}
\]  

(2.3)

First of all, the ILP problem (2.3) can be relaxed and solved as an LP problem using the simplex method. Therefore, the fractional optimal but infeasible solution \( z = 11.2, x_1 = 1.8, x_2 = 0.8 \) could be obtained. In order to make the solution feasible, as can be seen in Figure 2-1 we can choose to start from an integer variable \( x_i \) and branch the original problem \( z \) into sub-problem 1 with an extra added constraint \( x_i \leq 1 \) and sub-problem 2 with an extra added constraint, \( x_i \geq 2 \). Resolving the two sub-problems with simplex methods, one incumbent solution could be obtained from the left hand side of branch and bound tree with \( z = 14, x_1 = 1, x_2 = 2 \). On the other hands, on the right hand side of branch and bound tree we
obtain another fractional optimal solution \( z = 11.75, x_1 = 2, x_2 = 0.75 \). In similar idea, re-branch from variable \( x_2 \) by adding two extra constraints, we could obtain the third and fourth incumbent solutions. By comparison, the BB process finally terminated with the optimal solution: \( z = 13, x_1 = 2, x_2 = 1 \).

Figure 2-1. Example of Branch and Bound tree for ILP problem 2.3

However, BB methods also have its inherent shortcomings, as can be easily concluded from the aforementioned solving process, the size of branch and bound tree will dramatically increase when the scale or the number of integer variables in the MILP problem increases. For a realistic power system comprising thousands of generators, the MILP-UC model may contain a large number of binary variables. Thus, the resulting MILP problems can be computationally intensive for state-of-the-art implementations of BB methods.
In the following chapters, I will focus my study on the reformulation of the MILP-based UC problem in order to improve the computational speed of UC problem.

2.2 Non-Deterministic Optimization Methods

2.2.1 Introduction

Non-deterministic optimization is a serious of approaches for modeling optimization problems that involve uncertainty. Compared with the deterministic optimization problems, in which all parameters bear known values, the many parameters, such as forecast price, forecast load, are inevitably unknown at the time when decision should be made in the operation and planning study. When the parameters are uncertain, but assumed to be within certain range or a set of possible values, one might seek a solution that is feasible for all possible choices (or almost all) and optimize the objective function based on best average performance. Examples would be scheduling the generator on/off status and power output to minimize the overall operation cost based on the forecasted load or determining the energy storage charging/discharging status to maximize the owner’s profit based on forecast energy prices. In these contexts, stochastic programming is closely related to decision analysis, stochastic control theory, Markov decision processes, and dynamic programming [23]. In this thesis, I focus mainly on finding solutions of Multi-Armed Bandit Problems (MABP) in the discrete time stochastic control process.

2.2.2 The Multi-Armed Bandit Problem

In probability theory, MABP [23, 24] are a class of sequential resource allocation problems for allocating one or more resources among independent reward-producing Markov decision processes. All processes will change state randomly depending on weather they have resources been allocated or not. In the simplest setup of the MABP problem, if a
resource has been allocated to a process in a given time slot, the status of the process will change following a Markov process and a status-related reward will be generated. The status of the process with no resource been allocated will remain in the same status and there will be no reward for the process. Thus, the objective of the MABP is to maximize the expected average reward over a finite or infinite time horizon. The MABP represent one’s conflict between taking actions which could yield immediate reward and taking actions whose benefit will come in some later time.

The computational complexity of dynamic decision problems involving uncertainty is well known and the MABP was considered computational intractable for a long time [25]. For instance, the state space of process $i$ is $X^i$ with $m$ state. Then, the state space of player $X^i$ for each time step $i$ from 1 to $n$ can have a size of $m^n$. A key breakthrough for solving the MABP was made by Gittins [26, 27]. Gittins showed that instead of solving the n-dimensional Markov decision process (MDP) within state-space, a simple optimal solution can be obtained by solving 1-dimensional problem, in which the optimal strategy is always to select the process with the largest Gittins’ index. The results lead to one of the simplest classes of models for dynamic resource allocation. The models have been widely applied in many engineering applications such as signal processing [28], communication systems [29, 30], internet search [31], etc.

After Gittins, Whittle proposed an index policy in [32, 33] for the extension of the MABP called restless multi-armed bandit problems (RMABP). RMABP allows unselected processes (processes with no resource allocated) to evolve in a different Markov process from the selected processes. The goal of Whittle’s work is to seek the optimal policy of
selecting processes from independent restless processes \((m < n)\) to maximize the average reward over an infinite time horizon.

### 2.3 Solution Methods of UC Problem

A key objective of power system operation is to schedule generation to supply the power system load at minimum cost. As electricity demand has a typical diurnal and weekly pattern, generation units need to be carefully scheduled to meet this fluctuating demand in the most economic manner. The scheduling optimization is known as the UC problem, which is extremely important for both daily operation scheduling and planning studies from short term (e.g., weekly or monthly) to long term (e.g., yearly). The UC problem aims to minimize the total operational cost subjected to a group of system and generator constraints. It is a mixed integer programming problem with binary variables representing generation units' on/off status, and therefore is in the class of NP-hard problems.

Because of the NP-hardness, the exact optimal solution of the large-scale UC problem is normally difficult to obtain. Many solution techniques have been proposed and developed over the past decades. Examples include Dynamic Programming [15], Lagrangian relaxation methods [12, 13], PL methods [11, 9], and MILP through the BB method [8, 9, 14]. Dynamic programming approaches have been around the longest. In these methods, the state space consists of the on/off status of generation units. Lagrangian relaxation methods seek appropriate coordination for generating feasible primal solutions while minimizing the duality gap. Such methods can provide fast solutions for large-scale systems, but they often encounter difficulties in obtaining feasible solutions. Heuristic PL methods commit and uncommit units in a certain priority order (typically based on units' operating and
startup/shutdown costs). The PL methods are simple, fast and provide feasible solutions, but cannot guarantee optimal solutions.

By approximating a generators' nonlinear production cost as a piece-wise linear function, the mixed integer UC problem can be linearized to become a MILP. MILP has become very popular for solving the UC problem over the past decade due to advances in MILP solvers. However, solving a large-scale MILP UC problem is very time consuming and may require run times of several days for a long term planning study. Therefore, in the first phase of my research, I devoted my effort to reformulate the MILP-based UC problem for improving the computation speed. Different from previous MILP models, the lower number of binary variables in the reformulated MILP-UC formulation yields a reduction of branches in the branch-and-bound tree, so that less computing time is required by available solvers for the same UC problem. The details are introduced in the next chapter.
CHAPTER 3  A Computationally Efficient MILP Formulation for Solving Regional Grid Scheduling Problem

This chapter describes the process of reformulating the MILP UC problems for dimension reduction to improve the solution speed of UC problem. The network model, simulation environments, and the reformulation process are introduced in this chapter.

3.1 Nomenclature

3.1.1 Parameters

\( CSU_g \)  Start-up cost of unit \( g \) ($)

\( CSD_g \)  Shut-down cost of unit \( g \) ($)

\( DT_g \)  Minimum down time of unit \( g \) (hour)

\( F_{g,n} \)  Upper limit of the fuel input of block \( n \) in the piece-wise linear production cost function of unit \( g \) (MMBtu)

\( FC_g \)  Fuel cost of unit \( g \ )/MMBtu)

\( HR_{g,n} \)  Heat rate of unit \( g \) in block \( n \) (MMBtu/MWh)

\( LD_{i,t} \)  Load in zone \( i \) during time period \( t \) (MW)

\( N_g \)  Number of blocks of the piecewise linear cost function of unit \( g \)

\( P_{g} \)  Minimum power output of unit \( g \) (MW)

\( \overline{P}_g \)  Maximum power output of unit \( g \) (MW)

\( P_{g,n} \)  Upper limit of the power output of block \( n \) in the piece-wise linear production cost function of unit \( g \) (MW)
\( RE_i \) System reserve requirement of zone \( i \) as a percentage of its zonal load

\( RU_g \) Ramp up limit of unit \( g \) (MW/h)

\( RD_g \) Ramp down limit of unit \( g \) (MW/h)

\( T \) Unit commitment time span (hour)

\( TP_{g,t} \) Ambient temperature of unit \( g \) during time period \( t \) (°F)

\( UT_g \) Minimum up time of unit \( g \) (hour)

\( ZP_{x,y} \) Maximum power transfer limit from zone \( x \) to zone \( y \) (MW)

\( \Delta T \) Time step size (an hour)

3.1.2 Variables

\( c_{g,t} \) Production cost of unit \( g \) during time period \( t \) ($)

\( sud_{g,t} \) Start-up/Shut-down cost of unit \( g \) during time period \( t \) ($)

\( su_{g,t} \) Binary variable that equals to 1 if unit \( g \) start up during time period \( t \) and 0 otherwise

\( sd_{g,t} \) Binary variable that equals to 1 if unit \( g \) start up during time period \( t \) and 0 otherwise

\( u_{g,t,n} \) Binary variable that equals to 1 if unit \( g \) in block \( n \) of piecewise linear cost function is online in \( t \) and 0 otherwise

\( u_{g,t} \) Binary variable that equals to 1 if unit \( g \) is online in \( t \) and 0 otherwise

\( zt_{x,y,t} \) Power transfer from zone \( x \) to \( y \) during time period \( t \) (MW)

\( \delta_{g,t,n} \) Power production in block \( n \) of piecewise linear cost function of unit \( g \) during time period \( t \) (MW)

3.2 Network Model and Simulation Environments

This subsection introduces the network models used for the performance evaluation and the simulation environment.
3.2.1 **Network Models**

The Eastern Interconnection (EIC) network, one of the two major AC electrical grids in North America, is used in this study. The EIC network model is generated from the PROMOD IV database NERC 9.0, which contains 4622 thermal generators in 39 interconnected zones, as shown in Figure 3-1. PROMOD IV (http://www.ventyx.com/en/solutions/business-operations/business-products/promod-iv) is a commercial production cost model owned by Ventyx [34].

3.2.2 **Simulation Environments**

EOM is running on Ubuntu 14.04 Linux system on an Intel Xeon E5630 computer with 2.53GHz clock frequency and 4GB RAM. PROMOD IV is running on the same computer under Windows 7 Enterprise operating system.

Figure 3-1. Footprint of Eastern Interconnection.
3.3 **MILP Reformulation of UC problem**

The UC problem seeks to find the lowest cost schedule to commit/uncommit a fleet of generation units over an operation planning horizon to meet the demand, while adhering to system and individual generator constraints. Detailed formulation of the MILP-UC problem can be found in [8, 35]. For each generator, the operational constraints include capacity limitation, minimum up/down time, and ramp up/down. The decision variables for each generator include generator on/off status and generation output level, and startup/shutdown cost. For the whole power system, there are intra zonal transmission line constraints and system security reserve requirement constraints. The temporal interdependency and the combinatorial nature make large-scale MILP-UC problems computationally expensive to solve. This section will first summarize the scale of MILP-UC problem from given EIC database before the reformulation. Then, MILP-UC constraints will be reformulated to improve the computation speed.

Before the reformulation, using given network models and system parameters from database NERC 9.0, the MILP formulation of UC problem for EIC on hourly bases with 24 hours optimization time window can be seen from CPLEX log data as shown in Figure 3-2 and summarized below.

- Binary variables (integer variables): 387,907
- Continuous variables: 176,028
- Constraints: 624,782
Figure 3-2. Log data of formulated MILP-UC in CPLEX.

The UC problem with given scale take CPLEX around 3 hours to finish each optimization with 24 hours planning horizon under the simulation environment mentioned in 3.2.2. The slow solving speed is caused by the large number of binary variables in the UC problem. The reformulation of the MILP-UC problem will be introduced in the next few sections.

3.3.1 Reformulation of Start-Up Cost and Shut-Down Cost Function

The objective function of unit commitment problem can be written as

$$\min \sum_{t=1}^{T} \sum_{g} (c_{g,t} + CSU_{g}su_{g,t} + CSD_{g}sd_{g,t})$$ (3.1)

Then binary variables $su_{g,t}$ and $sd_{g,t}$ is bound by binary variable $b_{g,t}$, the on/off status of first block of piece-wise linear function of unit $g$ during time $t$. Equation 3.2 guarantees the binary variables $su_{g,t}$ and $sd_{g,t}$ equals to 1 if unit $g$ start up or shut down during time $t$. Equation 3.3 ensures that $su_{g,t}$ and $sd_{g,t}$ will not equals to 1 during the same time $t$. 

23
\[ su_{g,t} - sd_{g,t} = b_{g,t,1} - b_{g,t-1,1} \]  
(3.2)

\[ sd_{g,t} + su_{g,t} \leq 1 \]  
(3.3)

However, the binary variables \( su_{g,t} \) and \( sd_{g,t} \) with respect to each unit \( g \) at each time step \( t \) greatly increased the number of binary variables in the MILP formulation, and therefore greatly increased the scale and complexity of the define MILP problem.

The reformulated expression for object function and start-up shut-down cost function are as follows

\[
\min \sum_{t=1}^{T} \sum_{g=1}^{G} (c_{g,t} + sud_{g,t})
\]  
(3.4)

where \( sud_{g,t} \) is one continuous variable for unit \( g \) during time \( t \) to indicate start-up cost or shut-down cost. The start-up cost and shut-down cost of any unit could be combined into one continuous variable because no unit could start-up and shut-down simultaneously.

Then, the nonnegative continuous variable \( sud_{g,t} \) is bounded by binary variable \( u_{g,t} \), the on/off status of unit \( g \) during time \( t \), in order to indicate the start up or shut down actions. The constraints can be shown as

\[
sud_{g,t} \geq CSU_{g} (u_{g,t} - u_{g,t-1})
\]  
(3.5)

\[
sud_{g,t} \geq CSD_{g} (u_{g,t-1} - u_{g,t})
\]  
(3.6)

\[
sud_{g,t} \geq 0
\]

In the EIC power network with 24 hours optimization window, the reformulation of Start-up cost and Shut-down Cost function in 3.4-3.6 reduced \( 4622 \cdot 24 \cdot 2 = 221,856 \) integer variables at the cost of extra increased \( 4622 \cdot 24 = 110,928 \) continuous variables.
3.3.2 Reformulation of Piece-Wise Production Cost Function

In MILP formulation of UC problem, the quadratic production cost function of thermal generators is typically approximated by piecewise linear function. Such approximation is conceptual illustrated in Figure 3-3. As can be seen, the approximation error depends on the number of blocks in piece-wise linear function. If the blocks number \( n_r \to \infty \), the approximation error \( e \) also approaches zero \( (e \to 0) \).

![Figure 3-3. Conceptual illustration of piecewise linear approximation of production cost curve.](image)

The MILP formulation of piece-wise linear production cost function can be expressed as

\[
c_{g,t} = \sum_{n=1}^{n_r} \delta_{g,t,n} \Delta T \cdot FC_{g,t} \cdot HR_{g,t,n} \tag{3.7}
\]

\[
HR_{g,t,n} = \frac{F_{g,t,n} - F_{g,t,n-1}}{(P_{g,t,n} - P_{g,t,n-1})\Delta T}, \quad n \in [2, n_r]
\tag{3.8}
\]

\[
\delta_{g,t,n} \leq (P_{g,t,n} - P_{g,t,n-1})b_{g,t,n}, \quad n \in [2, n_r]
\tag{3.9}
\]

\[
\delta_{g,t,n} \geq 0
\]
\[ b_{g,t,n} \leq b_{g,t,n-1}, \quad n \in [2,n_g] \]
\[ 0 \leq b_{g,t,n} \leq 1 \]  \hspace{1cm} (3.10)

where \( \delta_{g,t,n} \) is power generation of unit \( g \) on block \( n \) during time \( t \), \( FC_g \) is fuel cost of unit \( g \), \( HR_{g,n} \) is heat rate (MMBtu/MWh) of unit \( g \) in block \( n \) during time period \( t \), which is a function of \( P_{g,n} \) (upper limit of the power output of block \( n \) unit \( g \)) and \( F_{g,n} \) (upper limit of the fuel input of block \( n \) of unit \( g \)). Binary variables \( b_{g,t,n} \) equals to 1 if unit \( g \) block \( n \) of piecewise linear cost function is online in \( t \). In addition, (3.10) is used to ensure that power output indicator \( b_{g,t,n} \) of unit \( g \) in block \( n \) will not be active when the previous block \( b_{g,t,n-1} \) is inactive \( b_{g,t,n-1} = 0 \).

The three dimensional binary variables \( b_{g,t,n} \) with respect to any thermal unit \( g \) in block \( n \) during time period \( t \) generated great amount of integer variables in MILP formulation, especially in a large scale MILP problem like EIC case. As we can observe from the quadratic production cost function in Figure 3-3, the heat rate of thermal generators increases with respect to the power output. Therefore, the block \( n \) of piece-wise linear function of any thermal unit \( g \) will not be active unless the previous block \( n-1 \) is active because the relatively higher heat rate at higher power output represents lower fuel cost efficiency at high power output compared with the fuel cost efficiency at low power output. Therefore, for each generators \( g \) during each time period \( t \), we only need one binary variable to indicate its on/off status.

Therefore, the piece-wise linear production cost function (3.9)-(3.10) can be simplified as
\[ 0 \leq \delta_{g,t,n} \leq (P_{g,n+1} - P_{g,n})u_{g,t} \quad (3.11) \]

In the EIC power network from database NERC 9.0, each thermal unit has three blocks piece-wise linear production cost function on average. The reformulation of Production Cost function in (3.11) has decreased \(4622 - 242 = 221.856\) integer variables.

### 3.3.3 Power Balance Constraints

Equation (3.12) guarantees the total power generation at each interconnected zone together with the power flow in \(p_i^{in}\) and power flow out \(p_i^{out}\) matches the zonal demand. \(z_{t,i,j}\) is power transfer from zone \(i\) to its interconnected zone \(j\) during time period \(t\). In addition, the power transfer between two inter-connected zones must satisfy its capacity limit in (3.13).

\[
\begin{align*}
\sum_{g=1}^{N_g} \sum_{n=1}^{N_n} \delta_{g,t,n} + p_i^{in} &= LD_{t,j} + p_i^{out} \\
p_i^{in} &= \sum_{j=1, j \neq i} z_{t,j,i} \\
p_i^{out} &= \sum_{j=1, j \neq i} z_{t,i,j} \\
z_{t,i,j} &\leq ZP_{i,j} \\
z_{t,j,i} &\leq ZP_{j,i} \\
\sum_{g=1}^{N_g} \sum_{n=1}^{N_n} \delta_{g,t,n} &= \sum_{i=1}^{I} LD_{t,j} 
\end{align*}
\quad (3.12, 3.13, 3.14) \]

However, the relation in constraints (3.14) will always be satisfied as long as constraints (3.12) are enforced. In addition, the redundant variables for power in \(p_i^{in}\) and power out \(p_i^{out}\) in each zone \(i\) zones are also removed. Therefore, the power balance constraints could be simplified as

\[
\sum_{g=1}^{N_g} \sum_{n=1}^{N_n} \delta_{g,t,n} + \sum_{j=1, j \neq i} z_{t,j,i} - \sum_{j=1, j \neq i} z_{t,i,j} = LD_{t,j} 
\quad (3.15)
\]
3.3.4 Minimum Up/Down Time Constraints and Ramp Up/Down Constraints

Minimum up/down constraints are ensured by (3.17) and (3.18), while ramping up/down constraints are ensured by (3.19) and (3.20),

\[
\sum_{k=1-UT_g+1}^{U} u_{g,k} \geq UT_g \left( u_{g,i} - u_{g,i+1} \right)
\]  
\[
\sum_{k=1-DT_g+1}^{D} (1-u_{g,k}) \geq DT_g \left( u_{g,i+1} - u_{g,i} \right)
\]  
\[
RU_g \geq \frac{\sum_{n=1}^{N_g} \delta_{g,i,n} - \sum_{n=1}^{N_g} \delta_{g,i-1,n}}{\Delta T}
\]  
\[
RD_g \geq \frac{\sum_{n=1}^{N_g} \delta_{g,i-1,n} - \sum_{n=1}^{N_g} \delta_{g,i,n}}{\Delta T}
\]

where \(UT_g\) and \(DT_g\) respectively represent the minimum up and down time of unit \(g\), \(RU_g\) and \(RD_g\) respectively represent the ramp up and down limit of unit \(g\).

3.3.5 System Reserve Constraints

The constraints for reserve requirement can be expressed as.

\[
\sum_{g=1}^{G} \left( \sum_{n=1}^{N_g} \delta_{g,i,n} - \sum_{n=1}^{N_g} \delta_{g,i-1,n} \right) \geq RE_i \cdot LD_{i,j}
\]

where \(\sum_{g=1}^{G} P_{g}^-\) represents the sum of thermal generator capacity in zone \(i\) and \(RE_i\) is the system reserve requirement of zone \(i\) as a percentage of its zonal load.

3.4 Simulation Results

This subsection presents the results comparison to validate the performance of the reformulation process.
3.4.1 Comparison of Problem Scale

The reformulated MILP-UC problem scale for EIC with 24 hours optimization time window could be observed from CPLEX log data as shown in Figure 3-4 and summarized below.

- Binary variables (integer variables): 387,907
- Continuous variables: 176,028
- Constraints: 624,782

![Figure 3-4. Log data of reformulated MILP-UC in CPLEX.](image)

Table 3-1. Comparison of MILP-UC problem scale before and after the reformulation.

<table>
<thead>
<tr>
<th></th>
<th>Before</th>
<th>After</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binary variables</td>
<td>387,907</td>
<td>110,928</td>
<td>28.6%</td>
</tr>
<tr>
<td>Continuous</td>
<td>176,028</td>
<td>214,971</td>
<td>122.1%</td>
</tr>
<tr>
<td>Constraints</td>
<td>624,782</td>
<td>426,796</td>
<td>68.3%</td>
</tr>
</tbody>
</table>

For the same UC problem, the comparison of problem scales before and after the reformulation is given in Table 3-1. From Table 3-1 we have following observations:

- The number of binary variables in the reformulated MILP accounts for only 1/3 of the binary variables in original problem. And all remain binary variables (110,928
=4,622*24) in the reformulated problem are \( u_{g,t} \), which represent the on/off discrete status of 4622 thermal units in EIC.

- The numbers of continuous variables increased 22.1% in the reformulated MILP because many discrete variables, such as \( su_{g,t} \) and \( sd_{g,t} \) in start-up/shut-down cost function \( b_{g,t,m} \) in the piece-wise linear production cost function, have been reformulated in continuous variables in MILP to increase calculation speed.

- The number of constraints have been reduced to 2/3 of original problem scale from reformulated MILP expression.

3.4.2 Comparison of Computation Speed

For the same UC problem, the comparison of running speed of between the reformulated MILP-UC problem and original MILP problem can be observed in Table 3-2. As can be seen, the running speed improved around 50 times after the reformulation in MILP problem format by using the given computer setup in section 3.2.2.

Table 3-2. Comparison of MILP-UC problem solution time before and after reformulation.

<table>
<thead>
<tr>
<th>Planning horizon time window</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 hours</td>
<td>3 hours</td>
<td>3.5 min</td>
</tr>
<tr>
<td>48 hours</td>
<td>&gt; 9 hours</td>
<td>11 min</td>
</tr>
</tbody>
</table>

* It is noted that the calculation speed includes data processing time of EOM and calculation time of CPLEX optimization engine.

3.4.3 Conclusions

The MILP-UC problem are very time consuming to solve for a long term planning study, especially for those large-scale power network like EIC. In this Chapter, the MILP-UC problem is reformulated to improve the computational speed of MILP problem formulation.
The reformulation is performed to each UC constraint as a first step to reduce the numbers of integer variables and continuous variables required for the problem formulation. After the problem reformulation, a number of integer variables can either be removed from the problem formulation or be transferred to continuous variables. The EIC network, with 4622 generators in 39 interconnected zones, is used to test the performance of the proposed MILP reformulation. The simulation results show significant improvement in computational speed after the MILP problem reformulation.
CHAPTER 4 A Two-Step Modified Priority List-Based MILP Method

In this Chapter, a two-step modified priority list-based MILP method is presented to further improve the solving speed of the UC problem.

4.1 Introduction

Previous research shows that 70%–85% of the constraints may be inactive when solving a UC problem [35, 36]. Removing those constraints at the beginning of the MILP solving process will not compromise the optimality. In a given power system that has stable demand patterns, the commitment status of most generation units in the generation fleet will remain unchanged for several days. Thus, the binary variables of those always on/off generators can be predetermined and removed from the MILP formulation to increase the solving speed of MILP-UC problem. Therefore, a two-step modified priority list (MPL) based MILP method is proposed to further improve the computational speed of UC programs. In the first step, a subset of units is committed or uncommitted according to a priority list which is generated based on their operating costs and demand. Scale factors can be used to adjust the amounts of committed and uncommitted generation. In the second step, the MILP problem is formulated with the corresponding generators’ on/off status removed from the set of binary decision variables, and solved by MILP solvers.

The rest of this chapter is organized as follows: Section 4.2 describes the impacts of different types of generation cost characterizations on UC problem results. Section 4.3
presents the method to combine MPL technique with MILP for the UC problem. Simulation results are presented and discussed in Section 4.4. Finally, Section 4.5 discusses conclusions.

4.2 Characterization of UC problem

![Pie chart showing percentage of generators by on/off status over the first week of January in 2010 for EIC system with 39 zones and 4622 generators.]

Figure 4-1. Percentage of generators by the on/off status over the first week of January in 2010 for EIC system with 39 zones and 4622 generators.

4.2.1 Capacity Characterization

Despite the complexity in the MILP-UC problem, there are certain patterns in UC results. Typically, many generators remain at the same commitment status over a time window of several days or even weeks. For example, from one week UC results from PROMOD IV, Figure 4-1 summarizes the percentage of generators by their on/off status over a week in 2010 for the Eastern Interconnection (EIC), modeled with 39 zones and 4622 thermal generators. As can be seen, about 66% of generators are never committed, while 28% are always online for the entire week. If the commitment status of at least some of those generators can be predetermined, a significant number of the binary variables can be removed from the MILP-UC problem, and the solution speed can be greatly improved.

Figure 4-2 illustrates the dynamics of load and generation during the week. The demand curve (in blue) in the same week is shown together with the reserve margin (in green) and total generation capacity (in red). As can be seen, almost all the base load is
served by the same set of generators (always “on” generators) over the period of one week. Meanwhile, there is a set of generators that are never committed nor used during that period (always “off” generators). From this observation, we can postulate that the base load and peak load levels or their scaled values may be used to determine the generation capacity that is either always “on” or always “off” during the planning period. Using this information together with the priority order of generation units, one can predetermine three sets of generators:

- Always “on” (committed) generators.
- Always “off” (uncommitted) generators.
- Generators with committed status to be determined.

![Figure 4-2. Demand curve vs. unit commitment results.](image)

4.2.2 **Generation Selection Methods**

The classifications of generators into one of the three sets can be done using the unit commitment costs, which include fuel cost ($/MMBtu), start-up/shut-down cost ($), variable
O&M ($/MWh) and fixed O&M ($/kW-yr). Here, I characterize and quantify the impacts of different types of production costs on UC results. Operating costs can be classified into two groups: 1) operation-related costs, including fuel cost ($/MMBtu) and variable O&M ($/MWh), and 2) nonoperation-related costs including startup and shutdown cost ($). Fixed O&M cost is excluded from the analysis because it exists as long as generators are in service and is decoupled from the UC results. The production costs for the same week in 2010 are summarized by category in Figure 4.3. As can be seen, the operation-related cost accounts for 75% of total cost. Startup and shutdown costs together account for less than 1% of total cost. Five cases are developed to further evaluate the relative impacts of different types of production costs on the objective function using EOM, as described below.

- Case 1: Base case.
- Case 2: 1 $/MMBtu increment in all generators’ fuel cost compared with base case.
- Case 3: 1 $/MWh increment in all generators’ variable O&M cost compared with base case.
- Case 4: $1000 increment in all generators’ startup cost compared with base case.
- Case 5: $1000 increment in all generators’ shutdown cost compared with base case.

![Figure 4-3. Production cost by category.](image)
Sensitivities were analyzed on the EIC system with 4622 thermal generators in 39 zones for a week in EOM. CPLEX is used as the optimization engine for MILP formulated in EOM. The relative MIP gap tolerance is set to be 0.1%. Figure 4-4 compares the total production cost for all five cases. As can be seen, case 2 results in the largest change in total production cost, followed by case 3, while cases 4–5 have almost no impact on total production cost. In other words, the total production cost is much more sensitive to operation-related costs than non-operation-related costs. Therefore, in this study, MPL is generated based on operation-related costs.

Figure 4-4. Sensitivity of production cost on different type of operational cost.

4.2.3 UC Problem Characterization

This subsection analyzes impacts of operation-related costs on generation UC status. The operation-related costs $C$ ($/\text{MWh}$) of each generator at its maximum power output is expressed in (4.1),

$$C = Fh + o$$  \hspace{1cm} (4.1)
where $F$ represents the fuel cost ($/MMBtu), $h$ represents the average heat rate (MMBtu/MWh, and $O$ is the variable O&M cost ($/MWh). The operation-related costs are calculated for each generator in the EIC. All the generators are ordered according to their operation-related costs from low to high. Figure 4-5 plots the cumulative generation capacity as a function of operation-related costs by category of commitment status. The blue region indicates the range of demand fluctuation in that week (292.6–391.3 GW). The red region indicates the range of operation-related cost variation of additional generators (besides the always on generators) required to support the fluctuating demand, without considering startup/shutdown cost and other operational constraints. The following observations can be made:

- Most generators with operation-related costs lower than the lower bound of the red region are always committed in the UC results.
- Few generators with operation-related costs higher than the upper bound of red region are always “on”.

![Figure 4-5. Cumulative generation capacity as a function of operation related costs by category of commitment status.](image-url)
Most generators with operation-related costs higher than the upper bound of red region are always uncommitted in the UC results.

Few generators with operation-related costs lower than the upper bound of red region are always “off”.

Generators with commitment status switched have operation-related costs of around the red region.

In summary, generator commitment results greatly depend on operation-related costs, which is consistent with the observation in the previous subsection.

4.3 Determination of Commitment Status

Based on the analysis of operation-related costs and commitment results, I propose a MPL-based MILP method for improving the computational speed of UC programs in this section. The flowchart of the proposed method is shown in Figure 4-6, where $D_{\text{min}}$ and $D_{\text{max}}$ represent the minimal and maximal demand in the planning period, respectively.
The proposed method involves two steps: 1) determination of on/off status for a subset of the units based on the demand curve and priority list, and 2) formulation of MILP for the remaining generation units. The first step includes two parallel threads: identification of always “on” generators and identification of always “off” generators. In the second step, the MILP problem is formulated and solved in EOM with the corresponding generators’ on/off status removed from the set of binary decision variables. In the first step, I order all \( N \)
generators in a list based on their operation-related costs. Then, based on the cumulative generation capacity of the ordered generator list and the range of demand fluctuation \([D_{\text{min}}, D_{\text{max}}]\), the range of operation-related costs \([C_1, C_2]\) of one additional generator required to support the fluctuating demand is calculated (like red region shown in Figure 4-5). Thus, the range of always “on” generators and always “off” generators in the ordered generators list can be identified based on two selected thresholds \(C_1 = F_1\) and \(C_2 = F_2\), where \(F_1\) and \(F_2\) are scale factors used to adjust the tradeoff between solution speed and level of optimality. \(P_g\) represent the maximal capacity of generator \(g\). To avoid over-committing the always “on” generators (or over-un-committing the always “off” generators), in the first step, the sum of maximum capacity of the committed always “on” generators in each zone is capped by the minimal zonal demand \(D_{\text{min}}(i)\) in the planning period, where \(D_{\text{min}}(i)\) and \(D_{\text{max}}(i)\) represent the minimal and maximal demand in the planning period of zone \(i\) in which generator \(g\) is located. Similarly, when determining the always off units, the summation of committable units’ maximum capacity \(P_{\text{max}}(i)\) in zone \(i\) is greater than the maximum of zonal demand \(D_{\text{max}}(i)\) in the planning period. Thus, when congestions occur in intra-zone transmission line, local generators in each zone can always supply the local demand. The optimization solver will seek for the most economical solution subjected to transmission line flow constraints.

### 4.4 Case Studies

A series of UC simulations over the first week of January in 2010 have been carried out, with an operation planning horizon equal to 24 hours. Performance of the proposed method is compared with different scale factors \(F_1\) and \(F_2\) in 4 cases:

- **Case 1:** Only identify always “off” generators in step 1.
• Case 2: Only identify always “on” generators in step 1.

• Case 3: Identify always “on” and always “off” generators (with selected scale factors $F_1 = F_2$).

• Case 4: Directly solve in EOM, no simplification in step 1.

• Case 5: $1000$ increment in all generators’ shutdown cost compared with base case.

The simulation results are shown in Figure 4-7, 4-8 and 4-9. Figure 4-7 compares the number of generators with commitment status predetermined in the first three cases. As can be seen, when scale factors $F_1 = F_2 = 1$, 59.6% generators’ on or off status are predetermined in the MILP problem in case 3, which means the scale of the MILP problem has been greatly simplified.

![Figure 4-7. Number of generators with commitment status predetermined in different cases.](image)

As can be seen, with different scale factors $F1$ and $F2$ in range $[1; 2.4]$, EOM solves the UC problem with similar optimality level in all four cases, with a difference less than 0.1% among all cases. This difference is mainly because of the relative MIP gap tolerance in CPLEX (0.1% in this study), rather than the nature of the problem.
Figure 4-8. Total production cost in different cases.

Figure 4-9 plots solution time vs. scale factor in different cases. When scale factors $F_1 = F_2 = 1$, the EOM solution speed in case 3 (proposed strategy) is 4.8 times faster than case 4 (no simplification) for a similar optimality level. As shown in Figure 4-9, the solution speed of case 1 is faster than the speed in case 2 for two reasons: 1) there are more predetermined generator numbers in case 1 than case 2; 2) In case 2, the generation output level variables associated with always “on” generators remain undetermined in the MILP problem, while all variables associated with always “off” generators have been removed from the MILP problem in case 1.

Figure 4-9. EOM solution time in different cases.
4.5 Conclusions

Solving the UC problem for several thousands of generators with full consideration of cost types and operational constraints is computationally expensive and often requires runtimes that are not acceptable for comprehensive scenario analyses used in exploring policy options. In this Chapter, based on pattern analysis of UC results and sensitivity of UC results to different costs, I presented a MPL-based MILP method that truncates the vast option space of the binary decision variables (always “on” generators and always “off” generators) to a smaller set of variables that can be efficiently solved in a MILP solver. The hybrid approach requires a two-step process to separate the generators. In the first step, a subset of units are committed or uncommitted based on the MPL and the demand curve. In the second step, the MILP problem is formulated with the corresponding generators’ on/off status removed from the set of binary decision variables and solved by MILP solvers. The simulation results show that the proposed method can significantly speed up the large-scale UC problem with negligible compromise in optimality.

Chapter 3 and 4 summarized my effort on improving the computational speed of the UC problem, as shown in Figure 4-10. In the next chapter, I will present the analysis conducted within the EOM framework: quantifying the impacts of heat waves on power grid operation.
Figure 4-10. The EOM modeling framework.
CHAPTER 5  Impact of Heat Waves on Power Grid Operation

This chapter presents the study of quantifying the impacts of extreme weather on power grid operation. This study also serves as a validation case study to compare the performance of the proposed MILP-UC problem reformulation and the modified priority list-based MILP method with the UC solution methods used in commercialized production tools.

5.1 Introduction

Since the start of 20th century, the average annual temperatures across United States have increased approximately 1.5°F (0.8°C) with 2001 to 2010 being the warmest decade ever recorded [37]. Global warming not only causes average temperature to rise but also leads to more frequent and severe extreme weather events such as heat waves [38] and droughts [39]. Quantifying the impacts of those extreme weather events on the reliability and operation of the power grid is becoming increasingly important, especially when large amounts of renewable resources and gas turbines have displaced coal power plants, which are less susceptible to variation of temperature, precipitation, and wind.

During the past a few years, many research efforts have been dedicated to study the impacts of climate change on electric power grids from different aspects. In [40], the benefits of power transfers between the Pacific Northwest and California were evaluated using a linked set of hydrologic, reservoir, and power demand simulation models for the Columbia River and the Sacramento-San Joaquin reservoir systems. In [41, 42], the impacts of climate change on electricity demand have been studied. The Building Energy Demand (BEND) model has been developed to simulate climate-dependent hourly building energy demands
with the ability to aggregate up to any geographic area, including counties, states, electric utility zones, or census regions [43]. An innovative platform called Regional Integrated Modeling and Analysis has been built to simulate interactions among natural and human systems at scales relevant to regional decision making [44]. In this chapter, I focus my effort on quantifying the climate change impacts on power system planning and production cost model brought by heat waves.

Heat waves could have significant impacts on power system operation, such as increased peak loads and reduced transmission and generation capacity. When a heat wave comes, the hot weather can last from days to weeks. Because more than 60% of the load is for cooling needs in summer, longer and hotter heat waves will cause sharp load increases [45]. In addition, the heat waves are usually accompanied by stationary high pressure zones, resulting in light winds at the surface. Wind generation will therefore decrease drastically. Moreover, higher ambient temperatures also reduce thermal capacity of transmission lines, which further stresses the power grid.

On generation side, a severe adverse impact from heat waves is the derating effects, i.e., reduced power output capacity and efficiency of generation units with combustion turbines. It has been known that turbine power output decreases and heat rate increases with increasing inlet air temperature [46-48]. Gas turbine (GT) and combine cycle gas turbine (CCGT) are the most important two types of gas fired plants in power generation fleet in the U.S. They are less pollutant, more efficient, and less costly energy sources compared with conventional coal plants [49]. Today, GT and CCGT combined together account for 27.4% of the nation’s annual electricity generation [50]. This number is expected to even increase in
the future [51], making the derating effects even more severe for reliable and economic operation of power systems during heat waves.

Figure 5-1. Illustration of impacts of heat waves on power system operation.

The combined impacts of a heat wave on grid operation is conceptually illustrated in Figure 5-1. As can be seen, the peak loads in this heat wave can be much higher than usual in both magnitude and duration. Meanwhile, the heat wave can also cause derating of GT and CCGT units potentially together with a reduction in wind generation. Such a deficiency in supply will result in severe power shortage problems in power system operation.

Production cost tools are used extensively in the electric power industry to forecast the expected amount of electricity produced by different power generation units and the expected cost of producing that electricity for a given system. The core of production cost analysis is UC that is used to determine which generation units should be committed over a short term scheduling period to meet the load and system reserve requirements, while adhering to generator operating constraints. Generator derating and load raise due to heat
waves can significantly affect unit commitment results and production cost analysis. However, existing studies relying on commercial production cost tools such as PROMOD [34] commonly do not reasonably model the load raise in heat waves. In addition, capacity of generators are defined on monthly basis, which is insufficient to reflect the consistently varying ambient temperatures. Moreover, the heat rate and efficiency of generators are assumed to be constant regardless varying temperatures. In order to quantify the potential impacts of heat waves in production cost analysis, it is desirable to model the load and generators based on hourly varying ambient temperature.

Literature review shows that although formulation of UC problems and the UC solution techniques have been well researched in the past [8, 9, 52, 53], the coupling between varying temperature on generation derating has not been considered yet. The main technical challenge of considering the derating of GT and CCGT units is that the generator output capacity and net efficiency will vary on an hourly basis. Because the numbers of GT and CCGT are significant in a large grid, such as the Eastern Interconnection, this temperature dependency requires significant data preprocessing of spatially and temporally distributed data to modify the hourly changing constraints and the cost functions in the UC problem. This Chapter will discuss how the hourly temperature profiles affect the UC formulation. As will be shown, the heat wave impacts can be addressed without increasing the dimension of the UC problem.

The rest of this chapter is organized as follows. In Section 5.2 and 5.3 hourly derating effects on GT and CCGT plants are characterized and incorporated into the formulation of the UC problem. Different scenarios are studied using EIC system in Section 5.4 to
demonstrate and quantify derating impacts from system security and economics perspectives. The conclusions are presented in Section 5.5.

5.2 Temperature Impacts on GT and CCGT Power Plants

This section will first characterize the derating impacts of GT and CCGT units due to increasing ambient temperature. These impacts are then incorporated into the UC formulation in production cost model.

5.2.1 Characterization of Temperature Impacts on Combustion Turbine

Combustion turbine is a dynamic internal combustion engine, which has a compressor coupled with a downstream turbine through a combustion chamber. Its power output capacity and efficiency depend on the ambient temperature, because any change in ambient temperature will affect inlet air flow mass, and consequently the power produced from a gas turbine.

Ambient temperature can vary dramatically at different geographic locations in different seasons in a large system. Taking EIC as an example, temperature in Texas frequently exceeds 95°F (35°C) in summer, while in Ontario, Canada, temperature can sometimes drop below -22°F (-30°C) in winter. As a result, the actual output capacity and net efficiency of combustion turbine at different location and time can be very different from its nameplate numbers. As surging temperature usually corresponds to peak electric power demand, the reduction in generation capacity during high ambient temperature is normally more problematic.

5.2.2 Capacity Derating

Capacity derating is typically modelled on monthly basis in commercial software packages, as illustrated in Figure 5-2. The generation capacity generally increases in winter
months when temperature is low and decreases in summer months when temperature is high. Such adjustment is capable to capture seasonal variation in temperature and its impacts on generation capacity. Nevertheless, the variation of temperature within a day can also be big, especially during heat waves.

![Graph showing monthly capacity from GT and CCGT in EIC in 2006.](image)

Figure 5-2. Monthly capacity from GT and CCGT in EIC in 2006.

In Figure 5-3, the temperature during a historical heat wave in Houston in 2000 is plotted together with the average value from 2000 to 2012 for a few days from August to September. As can be seen, during this heat wave, the highest temperature is 109°F (43°C) and the lowest is 73°F (23°C). The changes in generation capacity between day and night are significant. Therefore, using one capacity value for the entire month will lead to either an over- or under-estimation of available capacity for some hours.
The air mass contained in per unit volume decreases with increasing temperature. Less air mass results in less fuel mass can be ignited, and therefore reduced power output. Based on GT data sheets \[54\], the relationship between capacity derating rate \( \eta \) (unitless) and ambient temperature \( TP \) (in Fahrenheit) for GT can be expressed as

\[
\eta = \begin{cases} 
\frac{-0.3867TP + 122.54}{100} & \text{if } TP \geq 32^\circ F \\
1.10^{-\frac{TP - 32}{F}} & \text{if } TP < 32^\circ F 
\end{cases}
\]  

(5.1)

It is assumed that the operation of combustion mode accounts for 50\% of power output from a CCGT. Therefore, its capacity derating rate is half of GT. The derating rates for GT and CCGT as a function of ambient temperature are plotted in Figure 5-4.
Efficiency Derating

The combustion gas turbine burns an air-fuel mixture in combustion chamber to produce hot vapor for spinning the turbine. The density of input air decreases with increasing ambient temperature, resulting in more fuel needed to compress the same amount of air mass. The heat rate (MMBtu/MWh) is defined as

\[
\text{heat rate} = \frac{\text{fuel input}}{\text{power output}}
\]  

(5.2)

The increased fuel consumption per unit energy output leads to increased heat rate and decreased net efficiency of a GT and CCGT [55]. The normalized heat rate can be expressed as a function of ambient temperature, as shown in Figure 5-5.
5.3 MILP Formulation with Temperature Impacts

The detailed formulation of UC as MILP problem is presented in Chapter 3. With hourly varying ambient temperature, the output capacity and heat rate of GT and CCGT need to be adjusted. This will affect the objective function and some constraints in UC problem. I herein present the UC formulations affected by ambient temperature.

5.3.1 Generation Cost

\[
c_{g,t} = \sum_{n=2}^{N_c} \delta_{g,t,n} \Delta T \cdot FC_{g} \cdot HR_{g,t,n}^e
\]

(5.3)

\[
HR_{g,t,n} = \frac{F_{g,n} - F_{g,n-1}}{(P_{g,n} - P_{g,n-1}) \Delta T}
\]

(5.4)

\[
HR_{g,t,n}^e = \frac{HR_{g,n}}{\eta_{g,t}}
\]

(5.5)

where \(HR_{g,n}\) is nominal heat rate of unit \(g\) in block \(n\) during time period \(t\), which is a function of \(P_{g,n}\) (upper limit of the power output of block \(n\) unit \(g\)) and \(F_{g,n}\) (upper limit of the fuel input of block \(n\) of unit \(g\)). In order to capture the derating in heat rate during heat waves,
\( HR^* \) and \( 1/\eta^t \) are introduced to represent the derated heat rate and heat rate derating factor, respectively. In particular, \( 1/\eta^t \) can be calculated using the ambient temperature at time \( t \) according to Figure 5-5.

5.3.2 Generator Capacity Limits

The power generation limits of each unit are capped by its maximum power output and minimum power output

\[
P_t = \sum_{n=1}^{N_g} \delta_{g,n} \leq P_{g,t}^r
\]

(5.6)

\[
P_{g,t}^r = \eta_{g,t}^c \bar{P}_g
\]

(5.7)

where \( \bar{P}_g \) is the nominal capacity of unit \( g \). In order to capture capacity deating due to increasing temperature, \( \bar{P}_{g,t}^r \) and \( \eta_{g,t}^c \) are introduced to represent the derated capacity and capacity derating factor, respectively. In particular, \( \eta_{g,t}^c \) can be calculated using the ambient temperature at time \( t \) according to Figure 5-4.

5.4 Case Study

The derating model of combustion turbine together with the adapted UC formulation presented in previous section are used to quantify the heat impacts on power grid operation. In particular, EIC as the largest electric grid in North America is used for case studies in this study.
In this study, the temperature during normal summer period from weather stations in NLDAS (North American Land Data Assimilation System-2) [56] is used as baseline. A heat wave with 18°F (10°C) uplifts is then constructed for all EIC zones for time period between 7/19/2006 and 7/24/2006. The default zonal load in PROMOD IV database represents regular year, which is adjusted to reflect load change in the heat wave using BEND model (see [43] for details). Figure 5-6 plots the temperature and load in American Electric Power (AEP) Zone for both baseline and heat wave scenarios. Compared with a normal summer day, the average load increases by 31% during heat wave period.
5.4.1 **Derating Impacts on System Reserve**

The hourly generation capacity for all GT and CCGT plants are calculated for the temperature in baseline and heat wave scenarios. The hourly aggregated thermal capacity of GT and CCGT plants considering derating effects in EOM is compared with the monthly capacity used in PROMOD in Figure 5-7. As can be seen, in summer months from May to September, PROMOD underestimates the capacity by about 19 GW, which accounts for 9.0% of the total capacity from GT and CCGT. In winter months from October to April, PROMOD overestimates the capacity by about 3GW.
Figure 5-8. Derated capacity of CCGT and GT units in EIC from 7/15/2006 to 7/24/2006.

Figure 5-8 compares the derated capacity of CCGT and GT in normal summer days (7/15/2006-7/19/2006) and heat wave days (7/19/2006-7/24/2006) with nominal capacity. During normal summer days, the average capacity of CCGT and GT decreases 4.4 GW and 8.3 GW from its nominal value, respectively. During heat wave period, the capacity can decrease as much as 5.5 GW and 10.2 GW for CCGT and GT, respectively. This accounts for 4.6% and 9.5% their nominal capacities, respectively. The total capacity from CCGT and GT varies as much as 4 GW from day to night. Therefore, a fixed monthly derating factor cannot capture the impacts of varying capacity in UC and the associated production cost.
During heat waves, the reduced capacity of CCGT and GT together with the increased cooling loads leads to a sharp drop in system reserve. Figure 5-9 compares zonal reserve margin in normal summer days and heat wave days. The boxplot is a statistical data analysis tool that provides the statistical attributes such as mean (circle in box), median (thick bar inside the box), first and third quartile (bar outside the box), and potential outliers. As can be seen, the reserve is significantly reduced in almost all the zones due to heat waves. In normal summer days, only four zones are with an average reserve margin less than 30%. In heat wave days, there are 14 zones with an average margin less than 30%. The average zonal reserve margin across EIC decrease 22.3% from the normal summer days to heat wave days. For those zones with limited generation capacity, such as Zones 9, 15, and 21 (ISONEBOS, MINNESTA, NYZ-J), the derating effects could lead to severe power shortage problems during peak the demand hours. Therefore, insufficient consideration of heat wave impacts could lead to severe deficiency in power supply when a heat wave strikes the EIC power network.

Figure 5-9. Zonal reserve margin in regular summer days and heat wave days.
5.4.2 Derating Impacts on Production Cost

- Four cases are developed in order to evaluate derating in capacity and efficiency on production cost and LMP, as listed in Table 5-1.

- Case 1 serves as a base case, where capacity of generators is model on monthly basis, and heat rate is assumed to be constant regardless varying temperature. The study is carried out using PROMOD.

- Case 2, the capacity of GT and CCGT is modeled on hourly basis considering ambient temperature, while the efficiency is modeled the same as Case 1.

- In Case 3, heat rate of GT and CCGT is adjusted based on ambient temperature, while the output capacity is modeled the same as Case 1.

- In Case 4, derating effects in both capacity and efficiency are considered on hourly basis. This case captures the heat wave impacts most accurately in production cost analysis.

Table 5-1. Case definition.

<table>
<thead>
<tr>
<th>Cases</th>
<th>Software</th>
<th>Capacity derating</th>
<th>Efficiency derating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>PROMOD</td>
<td>monthly</td>
<td>considered</td>
</tr>
<tr>
<td>Case 2</td>
<td>EOM</td>
<td>hourly</td>
<td>not considered</td>
</tr>
<tr>
<td>Case 3</td>
<td>EOM</td>
<td>monthly</td>
<td>hourly</td>
</tr>
<tr>
<td>Case 4</td>
<td>EOM</td>
<td>hourly</td>
<td>hourly</td>
</tr>
</tbody>
</table>
Case 2 and 3 are first compared with Case 1 in order to quantify impacts of derating in capacity and efficiency separately. As an example, in Figure 5-10, LMPs in ALWSTTA (zone 2) are plotted for heat wave days. The differences in LMP occur mainly in peak load hours from 12pm to 3pm because 1) CCGT and GT are mainly used as peaking units due to their relatively high fuel cost compared with nuclear and coal plants in database NERC 9.0 in 2006, and 2) derating effects are most significant during these hours. The daily highest LMP in Case 1 is $9.6/MWh higher than Case 2 in average. This is because the monthly capacity in Case 1 underestimates the available capacity of CCGT and GT units during summer months. On the other hand, the daily highest LMP in Case 3 is $18.2/MWh higher than Case 1 due to the increased heat rate in Case 3. The boxplot of LMPs of all the zones within EIC in Case 1-3 are provided in Figure 5-11. The average LMP in Case 1 is 0.6% higher than Case 2 and 4.1% lower than Case 3.
Figure 5-11. LMP in EOM and PROMOD Case 1-3 during 7/12/2006~7/24/2006.

Figure 5-12. Production costs in Case 2-4 compared with Case 1 by generator category.

The difference in production cost by generator category between Case 2-4 and Case 1 are plotted in Figure 5-12. The following observation can be made.

- Case 2 has the lowest total production cost and production cost in CCGT and CT. This is consistent with the observation in LMP, as PROMOD underestimates the available capacity of CCGT and GT unit in summer months.
• Case 3 has the highest total production cost and production cost in CCGT and CT, because 1) the monthly capacity consistently underestimates the varying available capacity, and 2) the derated efficiency increases fuel usage.

• Case 4 represents the most realistic case, where hourly derating in both capacity and efficiency is captured. As expected, the total production cost and generation cost of CCGT and CT in Case 4 is higher than Case 2 but lower than Case 3 for aforementioned reasons.

• The difference in production cost in different cases is significant, which also indicates different unit commitment results and fuel usages.

In general, the degree to which the production costs are influenced by derating effects also depends on the fuel price of natural gas and the installed capacity of GT or CCGT. These are two key factors that determine the commitment results of GT and CCGT units. Because cost of natural gas per BTU is higher than nuclear and coal in NERC 9.0 in 2006, the derating impacts are mainly limited to the peak load hours. Nevertheless, as natural gas price declines, generation cost of GT and CCGT becomes more cost-effective in serving the base load. In addition, due to the economic, environmental, and technological changes in recent years, these gas fire plants become increasing popular. According to the Energy Information Administration, the annual energy generated from natural gas plants has increased 14% within 6 years, and their generation is expected to account for 80% of all added electricity generation capacity by 2035 [49]. Therefore, the impacts of derating on system reserve and production cost are expected to further increase in the years to come.
5.5 Conclusions

This chapter quantifies the potential impacts of extreme heat waves on UC and production cost analysis. The impacts of heat waves on capacity and efficiency of GT and CCGT are characterized. The derating effects are then incorporated into the UC formulation using hourly temperature profile. With the proposed approach, heat waves impacts are evaluated for EIC system. The results show that heat waves could have significant impacts on system reliability and production cost. Considering the derating impacts on an hourly basis is thus necessary to prevent: 1) overestimation of the derating impacts on the production cost in regular summer days, which may cause underuse of available generation capacity, and increase in production cost. 2) underestimation of derating impacts in heat wave periods, which may cause severe power shortages and large price spikes.
CHAPTER 6  A Real-Time Greedy-Index based PEV Dispatching Policy for Providing Ancillary Services

This chapter presents an extension of the power system resource scheduling and dispatch problems: the dispatch of distributed plug-in electric vehicles (PEVs) for providing grid ancillary services. The power shortage problems caused by extreme weather conditions include severe deficiency in power supply for meeting power system demand as well as flexible generation resources for providing ancillary services. On the other hand, the distributed controllable energy resources (such as PEVs, photovoltaic, etc) in modern electric power system environment provide potential capabilities for such power shortage problems. Traditionally, power system resource scheduling and dispatching problems are solved by using deterministic centralized management algorithms. Such as the MILP-UC methods I have introduced in Chapter 3, Chapter 4. However, as the number of controllable devices in distribution network (such PEVs, photovoltaic) rapidly increases over a wide range geographical area, the conventional centralized control algorithms might be difficult to maintain an optimal operating condition for so many devices in real-time [57-59]. Such real-time optimal control problems are further complicated by the uncertainties associated with those self-organized distributed resources (such as the arrival and departure time of PEVs). In this chapter, I extend my research from scheduling centralized generation resources to distributed energy resources.
6.1 Introduction

The adoption of plug-in electric vehicle (PEV) technology has made significant progress around the world in recent years. According to an EPRI report [60], the market penetration of PEV will be 37% by 2020 and 52% by 2035. The charging of PEV batteries represents a new type of the system load that can be shifted between a given time period for economic and grid operation considerations. In addition, PEVs equipped with vehicle-to-grid (V2G) technology [61], making it possible to discharge electricity to the grid whenever needed. Therefore, recent research has been focused on how to use PEV as resources for providing a wide range of grid services such as peak shaving and load shifting [62, 63], frequency regulation, and load balancing (LB) services [64-66] (e.g. frequency regulation and load following service).

Different from the energy services, high-value ancillary services have very strict requirements on the time, duration, and magnitude of the service provided. Because the control of the PEV charging and discharging is extremely fast and accurate, PEVs are ideal sources for providing such services. Authors in [64, 65] have compared the potential profit for the PEVs to provide grid regulations under the current payment mechanisms. Results show that a PEV battery rated at 10-15 kW can earn $3,777–5,000 per year by providing such services.

However, there are two main technical challenges for using PEVs to provide frequency regulation service: coordinative charging management and PEV battery lifetime degradation [67]. Compared with the magnitude requirement of frequency regulation service (normal 1~10 MW per hour), the power and energy capacity of a single PEV battery is small (normally 1.4~19.2 kW and 5~40 kWh based on SAE standard [68]). Therefore, hundreds or
thousands of PEVs are needed to satisfy the regulation service coordinatively. Moreover, any additional charging and discharging events will inevitably shorten the lifetime of the battery. The cost of the Li-ion battery is between $1000 and $1450 per kWh [69, 70], so it is essential to develop an optimal dispatching algorithm that can coordinate the charging and discharging of large number of PEVs to provide grid services with minimum impacts on battery lifetimes.

There are three key considerations for developing a real-time dispatch algorithm. First, the dispatch period of a PEV needs to be flexible to account for uncertainties in its arrival state-of-discharge (SOC) and time of departure. Second, the dispatching algorithm needs to be solved within one second so that the dispatch command can be sent to each PEV to follow the frequency regulation service signals on a 4-second basis. Thirdly, the PEV should be rewarded based on its contribution to the overall services. Furthermore, any additional lifetime degradation for providing such services should be compensated.

In [71], a stochastic optimization approach has been used for using PEVs to provide regulation and spinning reserves, in which uncertainties in PEV charging behaviors are modeled by probability distributions. A robust optimization model has been used in [72] to formulate the PEV management/coordination problem as a deterministic mixed-integer quadratic programming problem to handle the uncertainty. A fuzzy optimization approach is used in [73] to model the uncertainties in the ancillary service market and PEVs arrival/departure times. Their results show slightly improvement in the average SOC of PEVs. However, when solving large-scale PEV dispatching problems every few seconds over a period of a few hours, the number of stochastic states will soon increase to a level that the aforementioned optimization methods are computationally intractable or too slow to meet
the real-time dispatching needs. Furthermore, the solution optimality cannot be guaranteed when multiple scenarios are used in the problem formulation.

In [74], the authors proposed a real-time PEV charging management framework by using a three-level hierarchical structure to avoid solving a complex optimization problem directly. Under the proposed three-level charging structure, the complexity of PEV charging control problem can be significantly simplified and the results can be updated in real-time. In [75], the authors developed a Welfare Maximizing Regulation Allocation algorithm to coordinate PEVs for providing real-time regulation services. However, the framework in [74] does not consider the optimality of the solution and the algorithm in [75] can achieve asymptotically optimal performance only when the battery capacities are infinite.

In this Chapter, I proposed a real-time, greedy-index based dispatching policy (GIDP) to tackle the two main challenges: 1) solving difficulties caused by high dimensionality of the input variables and uncertainties, and 2) obtaining the global optimal solutions. Upon receiving the regulation signal at each decision time step, the GIDP will select the PEVs for charging, idling or discharging based on a priority list determined by a pre-calculated index. The index is associated with an individual PEV’s real time SOC and represents the cost for recruiting the PEV to provide regulation service considering three elements: payment for its service (proportional to the amount of service they provide), payment for the battery lifetime depreciation, and payment for delayed charging (a comfort measure). The global optimality of the solution from GIDP can be guaranteed by the mathematical proved “indexability”. The GIDP transforms the optimization problem from a high-dimensional space to a one-dimensional space so that the problem can be solved in real time. By solving the transformed problem in the one-dimensional space in real-time repetitively, the obtained solutions could
be inversely transformed back to the originally high-dimensional space and proved to be the global optimal for both spaces.

The connection of the developed PEV aggregator model with the rest of the EOM components is shown in Figure 6-1. The main goal is to allow the distributed energy resources be aggregated by a service provider so that they can be used to participate high-value ancillary service at the regional grid level. To achieve this goal, the aggregated service must meet the same service requirements that are applied to the generation resources. Therefore, in Section 6.2, the setup of the PEV aggregator service, the assumptions made regarding the communication and control capabilities, and the optimization problem formulation are first introduced. Then, the proposed greedy-index based dispatching policy is presented in Section 6.3. The simulation results are presented in Section 6.4 and Section 6.5. In Section 6.6, the research findings are summarized and the conclusions are presented.
6.2 Optimization Problem Formulation

The operation mechanism of the PEV charging system, the PEV charging system operation mechanism, PEV battery degradation model, and the complete optimization problem formulations will be introduced in this section.

6.2.1 Operation Mechanism of an Aggregator-Controlled PEV Charging System

The minimum bidding block for providing regulation services in a wholesale electricity market (usually more than 1MW for an hour) is usually much higher than the power and energy capacity of a PEV battery bank (normally in the range of 1.4~19.2 kW and 5~40 kWh). Therefore, given the large number of PEVs required for charging coordination, it is logical to assume that there is an aggregator who will negotiate with the Independent System Operator (ISO) on rates and service requirements and recruit the required amount of PEVs for providing the target services.
The following assumptions are made to establish the control framework of an aggregator-controlled PEV charging system (See Figure 6-2) for providing regulation services:

- Constant current and constant voltage (CC-CV) charging method is used so that both charging and discharging are at the rated power, \( p_i \).
- There are two-way communication links between the aggregator and each PEV. Upon arrival, each PEV will report its charging requirements to the aggregator. The information made available to the aggregator after the \( i^{th} \) PEV’s arrival includes: SOC state upon arrival \( s_i' \) and the expected SOC state at departure \( s_i'' \). The aggregator has total control of the PEV charging process.

![Figure 6-2. Configuration of an aggregator-controlled PEV charging system for providing regulation services.](image)

6.2.2 **PEV Action Options**

If the dispatching period is \([0,T]\) and the time step is \( \Delta t \), we have \( n_r = T / \Delta t \) decision making stages. At stage \( j \), the \( i^{th} \) battery can take one of the three actions: charging (1), idling (0), or discharging (-1). Thus, policy set \( U \) for a Markov Decision Process of \( n_{PEV} \) during service time \([0,T]\) can be defined as
Thus, for any PEV $i$ at stage $j$, the power output $p_{i,j}$ from power network side is

$$p_{i,j} = \begin{cases} 
-\eta_i^d p_i & \text{if } u_{i,j} = -1 \\
\frac{p_i}{\eta_i^c} & \text{if } u_{i,j} = 1 \\
0 & \text{if } u_{i,j} = 0
\end{cases}$$

where $\eta_i^c$ and $\eta_i^d$ is the one way charging and discharging efficiency of battery $i$.

### 6.2.3 SOC States of the PEV Batteries

Let's $s_{i,j}$ be the SOC state of the $i^{th}$ PEV. When $s_{i,j} = 1$, $soc_i = 0\%$, and $s_{i,j} = m_i$ when $soc_i = 100\%$. Let $e_i$ be the maximum capacity of the PEV battery. The maximum number of the SOC states $m_i$, can be calculated by

$$m_i = \frac{3600e_i}{p_i \Delta t}$$  \hspace{1cm} (6.3)

where $\Delta t$ is the decision making interval.

Thus, for $s_{i,j} \in [0, m_i]$, the corresponding $soc_i$ of the $i^{th}$ PEV can be calculated by

$$soc_i = \frac{s_{i,j} p_i \Delta t}{3600e_i}$$  \hspace{1cm} (6.4)

After each decision stage $j$, if PEV $i$ made selection between charging ($u_{i,j} = 1$), idling ($u_{i,j} = 0$) and discharging ($u_{i,j} = -1$), the SOC state of PEV $i$ will advance accordingly by
6.2.4 Frequency Regulation Services

The aggregator will set an hourly baseline load profile so it can provide the frequency regulation service by varying its charging loads above or below the baseline. It is assumed the aggregator will receive the frequency regulation signal from the ISO every 4 seconds, which is the interval of the California Independent System Operators’ AGC Automatic Generation Control signal at CAISO.

At the ancillary service market, regulation services can be separated into up and down services [69]. In this study, I formulate the optimization problem assuming that the aggregator provides one-directional service (regulation-up). This is mainly for illustration to make the value of charging, idling, and discharging easily distinguishable.

When providing the regulation-up service, the baseline of the aggregated PEV load is

\[ s_{i,j+1} = s_{i,j} + u_{i,j} \]  \hspace{1cm} (6.5)

\[ p_{\text{baseline}} = \sum_{i=1}^{n_{\text{pw}}} \frac{p_{i}}{\eta_i} \]  \hspace{1cm} (6.6)

So that the default status of a PEV at any decision stage \( j \) is “charging”. Then, at the beginning of the decision stage \( j \), the aggregator will determine which PEV(s) will be switched into idling or discharging to meet the regulation-up signal to minimize the overall associated costs is minimized.

The maximum regulation capacity is determined by \( p_{\text{baseline}} \). For example, assuming that an aggregator recruits 1000 charging PEVs with V2G capability and each PEV is rated at 10 kW, \( p_{\text{baseline}} \) is 10 MW. Then, the aggregator can provide up to 20 MW of regulation-up service.
Note that for the same 1000 controllable PEVs, we can also set $p_{\text{baseline}}$ to be 5 MW. Then, the PEVs can provide up to 15 MW regulation-up and 5 MW of regulation-down service. If regulation-up and regulation-down services are paid at the same price, the PEV aggregator will receive the same income whether it is providing 20 MW regulation-down or providing 10 MW bidirectional services. Therefore, although I use providing one-directional regulation-up services to develop the GIDP, the method can easily be extended to cases where the aggregator needs to provide bidirectional frequency regulation services.

6.2.5 **Aggregator Income and Rewards for PEVs**

An aggregator’s objective is to maximize his profits while providing sufficient incentives for PEVs to participate. An aggregator’s income depends on the committed capacity, $P_{\text{RegUp}}$, and the price of the committed service, $p_{\text{DA}}^{\text{RegUp}}$ at the day-ahead (DA) market. The aggregator’s income $inc$ for providing regulation service over the committed period, $[0,T]$ can be calculated as

$$inc = \sum_{0}^{T} P_{\text{RegUp}} \times p_{\text{DA}}^{\text{RegUp}}$$

(6.7)

Therefore, $inc$ is considered as a known value in the cost function for the real-time dispatch problem.

For example, let’s assume an aggregator has committed at the DA market to provide 1-MW regulation service at $50/MW from 7:00 a.m. to 7:00 p.m. Then, at 7:00 a.m. the next day, when the aggregator starts to dispatch the charging and discharging of PEV, the aggregator would have known his or her income after fulfilling the service requirement. Therefore, the problem can be formulated as that of minimizing the aggregator’s costs while meeting the regulation service requirement over the committed period, $[0,T]$.  

73
I assume that aggregators will make the following payments to incentivize a PEV owner to participate the ancillary service: a reward based on the amount of service provided, battery life-time depreciation compensation, and delayed charging compensation. Note that no payment will be made to a PEV when it is charging because PEVs in charging state do not contribute to regulation service.

Let \( r_i(u_{i,j}) \) represents the aggregator income contributed by the \( i^{th} \) PEV at stage \( j \):

\[
\begin{cases}
p_i \frac{inc}{\eta_i \sum_{j=0}^{n^T} r_{j|p}} & \text{if } u_{i,j} = 0 \\
0 & \text{if } u_{i,j} = 1 \\
p_i \frac{inc}{\eta_i \sum_{j=0}^{n^T} r_{j|p}} + p_i \frac{inc \cdot \eta_i}{\sum_{j=0}^{n^T} r_{j|p}} & \text{if } u_{i,j} = -1
\end{cases}
\]

(6.8)

where \( n^T \) is the number of decision intervals. Note that when in the “discharge” mode, a PEV provides almost twice as much regulation-up services as it does in “idling” mode because the PEV will not only provide consumption reduction by “not charging” but also inject extra power to the grid. However, this process results in additional battery lifetime depreciation. Therefore, a compensation, \( c_{i|p} \), will be made to PEV \( i \). In the next Section, I will discuss the battery modelling used in this study and the calculation of battery lifetime degradation cost.

### 6.2.6 Modelling of Battery Lifetime Degradation Cost

The lifetime of a battery can be estimated based on the number of charging cycles completed at different depth of discharge (DOD) [76-79]. For example, as shown in Figure 6-3(a), a NaS battery at DOD=50% and DOD=100% can cycle approximately 9525 and 3142 times, respectively. Figure 6-3(b) illustrates how a life cycle is counted when the DOD is at 50% and 100%, respectively. As can be concluded from Figure 6-3, the life of battery
depreciates faster at deeper DODs. This is because the battery cell resistance will increase at a lower SOC and the increased heat will trigger extra irreversible detrimental parasitic chemical reactions. Therefore, in my study, I assume that the lifetime degradation cost monotonically decreases with respect to the SOC of the PEV battery.

![Graph showing cycles versus depth of discharge and battery energy level over time for DOD 100% and 50%](image)

Figure 6-3. Number of charging/discharging cycles versus DOD and (b) Charging and discharging cycles at DOD=50% and DOD=100% [79].

Based on the counted battery life cycle at different DOD, battery life depreciations for the complete cycles, $L_{DOD}^{Complete}$ at each DOD are calculated as

$$L_{DOD}^{Complete} = n_{DOD}^{Complete} \frac{1}{C_{DOD}^{max}}$$  \hspace{1cm} (6.9)

However, in real operation, a PEV battery will switch between the charging and discharging modes frequently instead of completing the full charging/discharging cycles start from 100% SOC as shown in Figure 6-3. Therefore, I use the proposed equivalent charge
cycle model [76] to calculate the battery lifetime degradation cost from incomplete cycles.

\[ n_{\text{InComplete}} \]

6.2.7 Equivalent Charge Cycle Model

The cumulated residual heat from previous charging cycles can accelerate the battery degradation rate in subsequent charging cycles, especially when a battery frequently shifts between charging and discharging mode at low SOC.

Assume that the lifetime depreciations of an incomplete cycle equals to the lifetime depreciations difference between two complete cycles. Then, the battery life depreciation for the incomplete cycles, at each DOD can be calculated as

\[ L_{\text{Incomplete}} = n_{\text{InComplete}} \left( \frac{1}{C_{\text{StartDOD}}} - \frac{1}{C_{\text{EndDOD}}} \right) \]  \hspace{1cm} (6.10)

where \( C_{\text{StartDOD}} \) and \( C_{\text{EndDOD}} \) are the start and end DODs for the uncompleted cycle.

After that, the total life depreciation, \( L_{\text{EquivalentTotal}} \) can be calculated by summing up the life depreciations of all the complete and incomplete cycles.

The conceptual illustration of the equivalent charge cycles calculation for ESS life estimation can be seen in Figure 6-4.

Assumptions: \( C_{\text{Max 10\%}} = 5000 \), \( C_{\text{Max 20\%}} = 1000 \)

Equivalent lifetime calculation for uncompleted cycles

\[ L_{\text{Incomplete 10\% - 20\%}} = L_{\text{Complete 20\%}} - L_{\text{Complete 10\%}} = \frac{1}{C_{\text{Max 20\%}}} - \frac{1}{C_{\text{Max 10\%}}} = \frac{1}{1000} - \frac{1}{5000} = \frac{4}{5000} \]

Figure 6-4. Steps of calculating the equivalent charge cycles.
6.2.8 Compensation for Battery Lifetime Degradation Cost

The battery degradation cost of each incomplete charging cycle can be estimated by using the equivalent charge cycle estimation methods in Figure 6-4. However, the compensation of life-time depreciation of discharging action $c_{i,j}^{Life}(s_{i,j})$ does not equals to the degradation cost of each single discharging action $c_{i,j}^{Depreciation}(s_{i,j})$, because the cumulative degradation impact of this discharging action should be counted in addition to the direct degradation cost during $\Delta t$. As can be shown in Figure 6-5, for any PEV $i$ with battery charging curve in plan A, if we change the selection of PEV $i$ at any step $j^*$ (change idling into discharging). Then, we generate charging curve $A^*$. Therefore, the difference in the battery lifetime degradation cost of PEV $i$ between plan A and plan $A^*$ should be the degradation cost during step $j^*$ plus the difference in the battery degradation cost between plan A and plan $A^*$ in all subsequent time periods after step $j^*$. Expressing the battery SOC in plan A and plan $A^*$ as $s_{i,j}^A$ and $s_{i,j}^{A*}$, then we have

$$c_{i,j}^{Life}(s_{i,j}) = c_{i,j}^{Depreciation}(s_{i,j}) + \sum_{j=j^*+1}^{n} \left[ c_{i,j}^{Depreciation}(s_{i,j}^{A*}) - c_{i,j}^{Depreciation}(s_{i,j}^A) \right]$$

(6.11)

where the battery lifetime degradation cost $c_{i,j}^{Depreciation}(s_{i,j}^A)$ and $c_{i,j}^{Depreciation}(s_{i,j}^{A*})$ can be calculated by using Equivalent Charge Cycle Model introduced in Section 6.2.7.
To simplify the analysis, the cost difference between any charging plan A and plan A* can be approximated by using the cost difference SOC state $s_{i,j}$ and $s_{i,j}+1$, then the compensation of battery life-time depreciation of discharging action can be expressed as

$$c_{i,j}^{Life}(s_{i,j}) = c_{i,j}^{Depreciation}(s_{i,j}) + (N_r - j^*) (c_{i,j}^{Depreciation}(s_{i,j} + 1) - C_{i,j}^{Depreciation}(s_{i,j}))$$  \hspace{1cm} (6.12)

It is noted that many battery lifetime estimation algorithms [76, 79, 80] can be used to calculate the compensation of battery life degradation cost. The equivalent charge cycle model [76] is used because it’s a simple algorithm can be used provide accurate battery lifetime degradation cost estimation by using DOD data (as shown in Figure 6-3) only. Whereas many other battery lifetime estimation algorithms [79, 80] are more complex methods, and require extra testing for parameters adjustment in their algorithms.

6.2.9 Compensation for Delayed-Charging

Let’s define each PEV’s incentive award $c_{i}^{Award}$ in proportion to the aggregator income $r_i(u_{i,j})$ as $c_{i}^{Award} = k \cdot r_i(u_{i,j})$, where $k$ is a profit sharing factor. If $k = 0.5$, the aggregator will distribute 50% of the income earned by this PEV back to its owner. Thus, by adjusting the values of $k$, the aggregator can adjust the rewards for each PEV.

If an aggregator interrupts the charging of a PEV by letting it idle or discharge while its SOC is still lower than the owner defined minimal departure SOC state, $s'_i$, the aggregator will pay a penalty, $c_{i}^{Penalty}(s_{i,j})$, to the owner for his comfort sacrifice. This is because the interruption of the charging causes a delay in service. The compensation $c_{i}^{Penalty}(s_{i,j})$ can be calculated by
\[ c_{i}^{\text{Penalty}} = \begin{cases} (s_i^c - s_{i,j}) \frac{g_i(i)}{10} + g_z(i) & \text{if } s_{i,j} < s_i^c \\ 0 & \text{o.w.} \end{cases} \]  

where \( g_i(i) (g_i(i) \geq 0) \) and \( g_z(i) \) represents a predetermined penalty factor for adjusting the amount of the payment. And the slope \(|g_i(i)|\)is designed to be smaller than the maximum absolute value of slope in degradation cost curve \(C_{\text{Life}}\) with respect to SOC states. Figure 6-6 shows the \(c_i^{\text{Penalty}}\) curve with respect to SOC states.

![Figure 6-6. Conceptual illustration of \(c_i^{\text{Penalty}}\) where \(s_i^c = 95\%\).](image)

In summary, the costs of the aggregator for providing service include three payments: rewards distributed to each PEV for its contribution, compensation made to the PEV whose lifetime is shortened, and comfort compensation for delayed service.

### 6.2.10 Objective Functions and Constraints

Combining the cost and rewards function determined by (6.1)-(6.13), we have
\[
Z = \inf_{u_{i,j} \in \mathcal{U}} \left\{ \sum_{j=0}^{n_{PEV}} \sum_{i=1}^{n_{EV}} \left[ c_{i}^{Life}(s_{i,j}, u_{i,j}) + c_{i}^{Penalty}(s_{i,j}, u_{i,j}) + c_{i}^{Award}(u_{i,j}) \right] \right\} \\
\text{s.t.} \quad \sum_{j=1}^{n_{PEV}} p_{i,j} = p_{baseline}(j) - r_{i,j}^{up}, \quad j \in [1, n_{T}] \\
s_{i,j+1} = s_{i,j} + u_{i,j}, \quad j \in [1, n_{T} - 1] \\
c_{i}^{Life} = \begin{cases} 
    c_{i}^{Life}(s_{i,j}) \text{ if } u_{i,j} = -1 \\
    0 \quad \text{o.w.}
\end{cases} \\
c_{i}^{Penalty} = \begin{cases} 
    0 \quad \text{if } u_{i,j} = 1 \\
    c_{i}^{Penalty}(s_{i,j}) \quad \text{o.w.}
\end{cases}
\]

\[
c_{i}^{Award} = \begin{cases} 
    \frac{kp_{i} \cdot \text{inc}}{\eta' \sum_{j=0}^{n_{T}} r_{i,j}^{up}} & \text{if } u_{i,j} = 0 \\
    0 & \text{if } u_{i,j} = 1 \\
    \frac{kp_{i} \cdot \text{inc}}{\eta' \sum_{j=0}^{n_{T}} r_{i,j}^{up}} + kp_{i} \cdot \eta'' \frac{\text{inc} \cdot \eta''}{\sum_{j=0}^{n_{T}} r_{i,j}^{up}} & \text{if } u_{i,j} = -1
\end{cases}
\]

\[ u_{i,j} \in \{1,0,-1\} \quad i \in [1, n_{T}] \quad j \in [1, n_{PEV}] \]

Directly solving (6.14) for a global optimal solution is extremely difficult for three reasons. Firstly, this problem is a nonlinear integer programming (NIP) problem, which has nonconvex solution sets bounded by the feasible region of integer variables. Finding a global optimum to a non-convex problem requires one to prove that a particular solution dominates all feasible points by arguments other than the derivative-based approaches in convex programming. Second, the high dimensionality of (6.14) causes a huge number of state variables in the solution set. Assuming that there are 1000 PEVs in the control group, each PEV battery has 5000 SOC charging states. Then, for a 6-hour dispatching period with 4-second decision making intervals, the overall number of integer variables would reach up to. Third, the magnitude of the frequency regulation signal is unpredictable. In real-time dispatch, past decisions made are no longer changeable. However, each decision made will
significantly affect subsequent decisions. Therefore, an optimal decision made at step \( j \) may quickly become suboptimal when looking back in time.

To address these difficulties, I reformulate the problem (6.14) in the next section so that it can be solved using the proposed GIDP.

### 6.3 Greedy-Index Dispatching Policy

#### 6.3.1 Reformulation of the Finite Horizon Markov Decision Process

In operation research, such a sequential decision-making process considering uncertainties is called the Finite Horizon Markov Decision Process (FHMDP) [23]. Compared with the dispatching period, which is usually several hours, the decision making interval, \( \Delta t \), is very short (i.e. 4-second in this case) such that \( n_r = T / \Delta t \to \infty \). Let the average of the regulation up signal be \( k \) so that \( E\{ r_{up}^i \} = k \). Then, (6.14) can be relaxed into an infinite horizon MDP (IH-MDP) problem as

\[
Z_2 = \inf_{s_{i,j} \in \Omega} \left\{ \lim_{n_r \to \infty} \frac{1}{n_r} \sum_{j=0}^{n_r} \sum_{i=1}^{n_{\text{sys}}} \left( c_i^{\text{Life}} (s_{i,j}, u_{i,j}) + c_i^{\text{Penalty}} (s_{i,j}, u_{i,j}) + c_i^{\text{Award}} (u_{i,j}) \right) \right\}
\]

\[
\text{s.t. } \sum_{j=0}^{n_r} \sum_{i=1}^{n_{\text{sys}}} p_{i,j} = P_{\text{mandatory}} (j) - k
\]

The relaxation eliminates the uncertainties in the regulation service signal by transforming (6.14) from minimizing the aggregator’s total costs in \([0,T]\) to minimizing the aggregator’s average cost over an infinite time horizon.

#### 6.3.2 Lagrange Relaxation for Distributed Decision Making

Applying the Lagrange method to the equality constraint of regulation service requirements, we have
\[
L(W) = \sup_{w \in \mathcal{W}} \left\{ \inf_{U \in U} \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} \left[ c_i^{\text{life}}(s_{i,j}, U_{i,j}) + c_i^{\text{penalty}}(s_{i,j}, U_{i,j}) + c_i^{\text{award}}(U_{i,j}) + \sum_{j=0}^{n-1} w_j \left( P_{i,j} - P_{\text{baseline}}(j) + k \right) \right] \right\}
\]

\[
\Rightarrow \quad L(W) = \sum_{i=1}^{n_{\mathcal{W}}} \sup_{w \in \mathcal{W}} \left\{ \inf_{U \in U} \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} \left[ c_i^{\text{life}}(s_{i,j}, U_{i,j}) + c_i^{\text{penalty}}(s_{i,j}, U_{i,j}) + c_i^{\text{award}}(U_{i,j}) + w_j P_{i,j} \right] - \sum_{j=0}^{n-1} w_j \left( P_{\text{baseline}}(j) - k \right) \right\}
\]

(6.16)

Then, (6.17) can be naturally decoupled to sets of independent sub-problems through Lagrange relaxation so that the total payment to each PEV can be minimized independently. For example, the \(i\)th sub-problem of (6.16) is represented as

\[
L_i(W) = \sup_{w \in \mathcal{W}} \left\{ \inf_{U \in U} \lim_{n \to \infty} \frac{1}{n} \sum_{j=0}^{n-1} \left[ c_i^{\text{life}}(s_{i,j}, U_{i,j}) + c_i^{\text{penalty}}(s_{i,j}, U_{i,j}) + c_i^{\text{award}}(U_{i,j}) + w_j P_{i,j} \right] \right\}
\]

(6.17)

In (6.16) and (6.17), the Lagrange multiplier, \(w_j\), represent the changing rate of the objective function corresponding to any change in equality constraint of relaxed regulation service requirement at step \(j\). The solution of (6.16) and (6.17) provides a lower bound for problem (6.15). Based on the strong Linear Programming (LP) duality, there exists a multiplier \(W^*\) such that \(L(W^*) \leq Z^*_2\). If \(W^* \neq \tilde{0}\), the LP complementary slackness will ensure that an optimal solution of \(L(W^*)\) must satisfy the equality constraint of regulation service requirement in (6.15) and therefore will also be optimal for \(Z_2\). Thus, each PEV policy can now be calculated independently on (6.17) while still satisfying the equality constraints in (6.15).

In operation research, the sub-problem represented in (6.17) is called the W-subsidy problem, in which the Lagrange multiplier represents a dummy “subsidy” to subsidize the
inactive selections of each sub-problem. In this study, the inactive selection of each PEV is “continue charging”. Then, by comparing with the cost/profit of the active selections (i.e. “stop charging and idling” and “stop charging start discharging”), each PEV can decide independently whether to participate in the regulation service. Meanwhile, the equality constraint of regulation service requirement is still satisfied by the appropriately selected. For more details regarding the W-subsidy problem, please refer to [32, 33].

Defining the participation cost (per kW) of using PEV $i$ for regulation service with SOC state as $\pi_i(s_{i,j}, u_{i,j})$, we have

$$\pi_i = \begin{cases} \frac{c_i^{penalty}(s_{i,j}, u_{i,j}) + c_i^{ Award}(0)}{p_i \eta_i} & \text{if } u_{i,j} = 0 \\ \frac{c_i^{penalty}(s_{i,j}, u_{i,j}) + c_i^{penalty}(s_{i,j}, u_{i,j}) + c_i^{ Award}(1)}{p_i + p_i \eta_i^d} & \text{if } u_{i,j} = -1 \end{cases} \quad (6.18)$$

It is noted that the power contribution in idling selection and discharging selection are different. A PEV provides more regulation-up services in “discharging” mode than it does in “idling” mode because the PEV will not only provide consumption reduction by “not charging” but also inject extra power to the grid.

At any step $j$, the selection for PEV $i$ depends on the value of $\pi_i(s_{i,j}, u_{i,j})$ and $w_{i,j}$ such that

If $\pi_i(s_{i,j}, u_{i,j} = 0) \leq \pi_i(s_{i,j}, u_{i,j} = -1) \& \pi_i(s_{i,j}, u_{i,j} = 0) < w_{i,j}$

Let PEV $i$ idle.

If $\pi_i(s_{i,j}, u_{i,j} = -1) < \min(\pi_i(s_{i,j}, u_{i,j} = 0), w_{i,j})$

Let PEV $i$ discharge.
If \( w_{i,j} \leq \min(\pi_i(s_{i,j}, u_{i,j} = 0), \pi_i(s_{i,j}, u_{i,j} = -1)) \)

Let PEV \( i \) remain charging.

This comparison process represents a **myopic policy** that makes the most profitable selection at each decision step. In the convex problem, the cumulated result obtained using such a **myopic policy** can be the same as the global optimal solution. However, the PEV dispatching problem is a non-convex problem so there is no guarantee that by selecting the optimal selection at each step, one will achieve the global optimality at the end of the dispatch period \( T \).

The theory of stochastic dynamic programming guarantees the existence of an optimal policy that is stationary (time invariant), deterministic, and Markovian. This is proved in [80] for the IH-MDP. In the next subsection, I will introduce the property of “indexability” proposed by Whittle [32, 33] to guarantee the existence of a unique global optimal solution in (14) by properly designing the cost and award functions at each SOC.

### 6.3.3 Greedy Index based Policy for Solving W-Subsidy Problems

In (6.18), it’s plainview that \( \pi_i(s_{i,j}, u_{i,j}) \) can be pre-calculated for each SOC state \( s_{i,j} \). However, \( w_{i,j} \) needs to be calculated at each step. In [32, 33], Whittle identifies a key property called “indexability” that is sufficient to guarantee the existence of a unique asymptotical optimal solution to the family of **W-subsidy problems** in (6.17). The indexability depends on the parameter \( \Pi(S,U) \) and \( W \) in a very simple fashion.

In policy set \( U \), let \( A_1 = \{ x \in U | x = -1 \land x = 0 \} \) be the collection of all active selections. Then, an active selection is either “stop charging and discharging” or “stop charging and
idling” (\( u_{i,j} \in \{0, -1\} \)). Similarity, let \( A_2 = \{x \in U \mid x = 1\} \) be the passive set. Then, a passive selection is “continue charging” (\( u_{i,j} = 1 \)).

**Definition 1** (Whittle [33]) A sub-problem is said to be indexable if the cardinality of set \( |A_i| \) monotonically increases from zero (empty set \( |A_i| = 0, A_i = \emptyset \)) to \( n_{PEV} n_T \) (full set \( |A_i| = n_{PEV} n_T, A_i = U \)) when the subsidy \( w \) for the passive set \( (A_2) \) increases from \(-\infty\) to \(+\infty\). If all sub-problem are indexable, the index-based policy is applicable to the W-subsidy problem.

**Definition 2** (Whittle [33]) Represent the index associated with PEV \( i \) at decision interval \( j \) as \( I(S) \), where \( i(s_{i,j}) \in I(S) \). If all the sub-problems are indexable, at any step \( j \), the PEVs with lower index are active under the W-subsidy policy.

**Definition 3** (Whittle [33]) If all the sub-problems are indexable, the index function \( i(s_{i,j}) = w_{i,j} i_i(s_{i,j}) \in I(S) \) of project (PEV) \( i \) in any state \( s_{i,j} > 0 \) is designed as the minimal value of the dummy subsidy that makes the two selections: participating in the regulation service (discharging or idling) and not participating in the regulation service (charging) equally profitable.

Based on Whittle’s definition in [33], the index \( i_i(s_{i,j}) \) represents the cost for an aggregator to activate PEV \( i \) with SOC state \( s_{i,j} \). If all \( n_T \) sub-problems defined in (6.17) can be proved as “indexable”, the optimal policy is “always activate PEVs that have relatively lower index (cost) at any decision step \( j \)”.

Following Definition 3, the index function \( i_i(s_{i,j}) \) can be chosen as the minimum of participation cost per kW, \( \pi_i(s_{i,j}, u_{i,j}) \). And \( i_i(s_{i,j}) \) can be defined as,
\[ i_j(s_{i,j}) = \begin{cases} 
\pi_i(s_{i,j},-1) & \text{if } \pi_i(s_{i,j},-1) \leq \pi_i(s_{i,j},0) \& \text{PEV } i \text{ with V2G} \\
\pi_i(s_{i,j},0) & \text{o.w}
\end{cases} \quad (6.19) \]

Let’s define PEV \( i \)'s active selection \( u_i \) with respect to state \( s_{i,j} \) as \( u_i(s_{i,j}) \). Then, we have

\[ u_i = \begin{cases} 
-1 & \text{if } \pi_i(s_{i,j},-1) \leq \pi_i(s_{i,j},0) \& \text{PEV } i \text{ with V2G} \\
0 & \text{o.w} \end{cases} \quad (6.20) \]

If all \( N_{\text{PEV}} \) sub-problems defined in (6.18) are proved as “indexable”, the resource dispatching policy can now be greatly simplified by following a three-step process that guarantees the optimal solution of (6.15) in the entire dispatching time horizon. The detailed steps of the proposed greedy-index based dispatching policy (GIDP) are described in Table 6-1. The prove on the “indexability” of all PEVs in (6.18) between SOC states \([s_i^2, s_i^3] \) is provided in next section, where \( s_i^2 < s_i^3 \) with \( \pi_i(s_i^2, -1) \geq \pi_i(s_i^2, 0) \).

**Table 6-1. The greedy-index based dispatching policy (GIDP).**

1. At each decision step \( j \), calculate index \( I(s) \) for each PEV.

   If PEV \( i \) with V2G capability,
   
   Calculate \( i_j(s_{i,j}) \) and \( u_i(i_j) \) based on (6.19) and (6.20).

   If PEV \( i \) without V2G capability,
   
   \[ i_j(s_{i,j}) = \pi_i(s_{i,j},0), u_i(i_j) = 0. \]

2. Sort the index in ascending order and use it as the priority list.

3. Proceed down the priority list and select PEVs for regulation service.

   If \( r^{UP}(j) \geq 0 \& u_i(i_j) = -1 \),
   
   \[ r^{UP}(j) = r^{UP}(j) - p(l \eta_i^u - p, \eta_i^d); \]
Switch PEV $i$ from charging to discharging.

Else if $r^{UP}(j) \geq 0$ & $u^*_i(i) = 0$,

$$r^{UP}(j) = r^{UP}(j) - p_i \eta_i^*;$$

Switch PEV $i$ from charging to idling.

Else

All the remaining PEVs will remain in charging states.

STOP

6.3.4 Prove of Indexability

Given the assumption of battery degradation costs $c_i^{Life}$ in section 6.2.6 and delayed-charging compensation $c_i^{Penalty}$ in Figure 6-6, both $c_i^{Life}$ and $c_i^{Penalty}$ monotonously decrease with respect to SOC state $s_{i,j}$ between $[1,s_i^c]$. Therefore, the participation cost per kW of using PEV $i$ for regulation service $\pi_i(s_{i,j},u_{i,j})$ the participation cost of PEV $i$ can be summarized as

$$\pi_i(1,u_{i,j}) > \pi_i(2,u_{i,j}) > ... > \pi_i(s_i^c,u_{i,j})$$

(6.21)

For any PEV $i$ with V2G capability, let’s define SOC state $s_i^c$ with $\pi_i(s_i^c,0) \geq \pi_i(s_i^c,-1)$ between $[1,s_i^c]$. Because the slope of $c_i^{Penalty}$ is designed to be smaller than the absolute value of maximum slope of $c_i^{Life}$. Therefore, $u_{i,j}(s_{i,j}) = 0$, if $s_{i,j} \in [1,s_i^c]$, and $u_{i,j}(s_{i,j}) = -1$ if $s_{i,j} \in [s_i^c,s_i^e]$.

For any PEV $i$ with V2G capability and SOC state $s_{i,j} \in [s_i^c,s_i^e]$, assume it has one fixed Lagrange multiplier $w_i$ with $w_i \geq \pi_j(s_i^e,u_{i,j})$. Based on the $W$-subsidy problem (6.18), no matter what initial SOC state PEV $i$ might start from, within finite time, the PEV $i$, will invariably
reach some state of $s_i^*$ with $s_i^* \in [s_i^2, s_i^e]$ and therefore alternating sojourns in state $s_i^*$ and $s_i^* - 1$

with average cost $\frac{\pi_i(s_i^*, -1) + w_i^*}{2}$.

Given final sojourn state $s_i^*$, the W-subsidy problem (6.18) can be rewritten as

$$L_i^*(w_i^*) = \sup_{-\infty < w_i < \infty} \left( \frac{\pi_i(s_i^*) + w_i^*}{2} \right) \quad (6.22)$$

In (6.22), if the Lagrange multiplier $w_i^*$ associated final sojourn state $s_i^*$ is expressed as

$w_i(s_i^*)$, then for all states between $s_i^2$ and $s_i^e$, we have

$$w_i(s_i^2) \geq \pi_i(s_i^2, -1) > w_i(s_i^2 + 1) \geq \pi_i(s_i^e) \quad (6.23)$$

If the final sojourn state $s_i^*$ with respect to any given Lagrange multiplier $w_i$ is expressed as $s_i^*(w_i)$, Given the relation in (6.23), it is straightforward to show that for any PEV $i$, there must be a one to one relationship between final sojourn state and Lagrange multiplier as $s_i^*(w_i(s_i^2)) = s_i^2$, $s_i^*(w_i(s_i^2 + 1)) = s_i^2 + 1, \ldots$ with relationship

$$s_i^*(w_i(s_i^2)) < s_i^*(w_i(s_i^2 + 1)) < \ldots < s_i^*(w(s_i^e)) = s_i^e \quad (6.24)$$

Therefore, for any PEV with SOC between $[s_i^2, s_i^e]$, the cardinality of passive set $A_i$ will increase monotonically from the empty set to the full set of all states as all passive dummy subsidy (cost) $w_{i,j}$ increases from $-\infty$ to $+\infty$.

End Prove.
6.4 Optimality Validation

When dispatching the operation of a large number of PEVs with many possible operation states and decision making stages, using enumeration or dynamic programming is computationally intensive and sometimes the problem is impossible to solve. For instance, for a PEV charging group consisting of 10 PEVs with 100 SOC states and vehicle-to-grid (V2G) functions, the number of possible states (charging, idling and discharging) at each single decision interval in dynamic programming is $10^{100}$. Therefore, if the forward dynamic programming method is used to obtain a global optimal solution for (6.14), the size of the problem has to be fairly small. Thus, in the first case study, a charging group consisting of 3 PEVs is used to illustrate the GIDP process and to compare its performance with dynamic programming.

When using forward dynamic programming to dispatch the PEVs, it is assumed that the regulation signal in the dispatching period $[0, T_n]$ is known so that the solution inaccuracy caused by regulation forecasting errors can be eliminated.

6.4.1 Operation Mechanism

The operation mechanism of the aggregator-controlled PEV charging System is illustrated as follows. After a vehicle arrives at the charging station, the vehicle owner will submit a charging request (see Table 6-2) to the aggregator.

Table 6-2. An example of the customer charging request.

<table>
<thead>
<tr>
<th>$s_i^a$</th>
<th>State of Charge (SOC) state upon arrival</th>
</tr>
</thead>
<tbody>
<tr>
<td>$s_i^f$</td>
<td>Expected SOC state at departure</td>
</tr>
<tr>
<td>$P_i^{rated}$</td>
<td>Rated power of the PEV battery (kW)</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Efficiency of the PEV battery</td>
</tr>
<tr>
<td>----------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td><strong>Vehicle-to-grid capability</strong></td>
<td>$1 \text{– Yes; } 0 \text{– No}$</td>
</tr>
<tr>
<td><strong>Sign-up for participating the regulation services</strong></td>
<td>$1 \text{– Yes. Provide grid service} \ 0 \text{– No. None-stop charging till } s_i' \text{ is met}$</td>
</tr>
</tbody>
</table>

Once the request is received, the aggregator can alter the PEV charging process for vehicles signed up to provide frequency regulation service. The aggregator can pay each PEV the following three kinds of payments: participation award, life-time depreciation compensation, and delayed charging penalty.

### 6.4.2 Simulation Setup

The parameters and charging requests are listed in Table 6-3. Note that the expected departure SOC state, $s_i'$, of each vehicle can be very different.

<table>
<thead>
<tr>
<th>PEV</th>
<th>Batter power and energy capacity</th>
<th>SOC at arrival $s_i'$</th>
<th>Expected SOC at departure $s_i'$</th>
<th>Penalty factors $g_1(i) \cdot g_2(i)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 kW/40 kWh</td>
<td>22.59 kW (56.475%)</td>
<td>39.15 kW (97.88%)</td>
<td>-0.004, 0.008</td>
</tr>
<tr>
<td>2</td>
<td>1 kW/40 kWh</td>
<td>9.30 kW (23.23%)</td>
<td>37.83 kW (94.57%)</td>
<td>-0.004, 0.008</td>
</tr>
<tr>
<td>3</td>
<td>1 kW/40 kWh</td>
<td>18.82 kW (47.05%)</td>
<td>35.32 kW (88.30%)</td>
<td>-0.004, 0.008</td>
</tr>
</tbody>
</table>
Figure 6-7. Comparison of different cost parameters of different PEVs.

The penalty and compensation for degradation cost curves are plotted in Figure 6-7. In Figure 6-7, the dotted lines are the cost curves for penalizing delayed charging. The penalty curves for all PEVs take the same penalty factors. However, once the expected SOC at departure of a PEV is met, its penalty cost will drop to zero. For example, for PEV 3 the expected departure SOC is 35.32 MW and above, so the penalty cost drops to zero once its SOC is above 88.30%. On the other hand, the solid lines represent the lifetime depreciation cost of each PEV. As shown in Figure 6-7, it is assumed the PEV 3 battery depreciates the fastest and the PEV 1 battery depreciates the slowest. It is straightforward that PEVs with lower departure SOC requirement and lower degradation cost are more competitive when providing the regulation services. Refer to Sections 6.2.6-6.2.9 for more details regarding the compensation for battery degradation cost and delayed charging penalty.
Based on equation (6.12) and (6.13), we can plot the cost curves for each action (charging, idling, or discharging) and the GIDP index curve with respect to the SOC. As shown in Figure 6-8, for example, the most profitable active selection of PEV 2 is: idling if \( SOC_2 < 19\% \) or \( SOC_2 > 94\% \); discharging if \( 19\% \leq SOC_2 < 94\% \). Recall that the corresponding GIDP index is calculated by (6.19). Then, for PEV 2, we have

If \( SOC_2 < 19\% \), then \( \pi_j(s_{2,j}, 0) \);

If \( 19\% \leq SOC_2 < 94\% \), then \( \pi_j(s_{2,j}) = \pi_i(s_{i,j}, -1) \);

If \( SOC_2 > 94\% \), then \( \pi_j(s_{2,j}) = \pi_i(s_{i,j}, 0) \).

After the index value \( I(S) \) is calculated for all three PEVs, we can then select PEVs for regulation service by comparing their index values. For example, at decision step \( j \), we need 2 kW regulation-up service, assuming that the SOC of each vehicle is 30%, we have
PEV 1: $i_{s_{1j}}(s_{1j}) = 0.0013$, $u_{i_{1j}}(i_{1j}) = -1$;

PEV 2: $i_{s_{2j}}(s_{2j}) = 3.72 \times 10^{-04}$, $u_{s_{2j}}(i_{2j}) = -1$;

PEV 3: $i_{s_{3j}}(s_{3j}) = 0.0021$, $u_{s_{3j}}(i_{3j}) = 0$.

As shown in Table 6-1, we should select PEVs to meet the regulation signal based on their index value in an ascending order. Therefore, PEV 2 (from charging to discharging) is selected to provide the 2 kW regulation-up service.

6.4.3 Performance Comparison

![Regulation signal over a period of 10 minute.](image)

Figure 6-9. Regulation signal over a period of 10 minute.

A 4 kW frequency regulation signal (see Figure 6-9) was constructed for Case 1. The PEV charging trajectories are computed first using the proposed GIDP method and then using the forward dynamic programming method. The results are plotted in Figure 6-10. The following observations are made:

- In the first half hour, because the SOCs of the PEVs 2 and 3 are low, both $C_{Life}$ and $C_{Penalty}$ are high on PEV 2 and PEV 3. On the other hand, PEV 1 has a relatively low degradation
cost and relatively high SOC. Therefore, discharging PEV 1 is more profitable than using PEV 2 or PEV 3 in the first half hour.

- When the SOC of PEV 1 decreases, discharging becomes a costlier option because $c_{i,j}^{\text{L}}$ is significantly higher at lower SOCs. For example, in hour 2, when the SOC of PEV1 is below 24 kW/h, Selecting PEV1 for regulation service is no longer more profitable than selecting PEV 2 or PEV 3.

- All three vehicles can meet the expected departure SOC requirements at the end of the dispatching period because the penalty mechanism tends to let the vehicles with greater unmet SOCs to charge first.

Because all PEV’s active selection $u_{i,j}(s_{i,j})$ remain unchanged on most SOC states (see Figure 6-8) and the charging decision were made based on the PEV GIDP indexes, the charging trajectory of each PEV is smooth with very little “zigzag” behaviors.

![Figure 6-10. Comparison of the charging trajectories of proposed GIDP method.](image-url)
The cumulated profits of the two approaches are shown in Figure 6-11. The two curves are fairly close to each other toward the end of the charging process. However, the profit curve from GIDP has a minor deviation from that of the dynamic programming in the middle of the charging process for the following reasons:

- GIDP seeking the optimal policy to maximize the average profit when $N_T \to \infty$. The optimal solution of maximizing average profit over $[0,T]$ may not be exactly the same optimal solution when $N_T \to \infty$.

- The reformulation relaxation in (6.15) and the Lagrange relaxation (6.16) relaxed the original problem in (6.14) and generate the optimal solution as $L(W') \geq Z'$. Therefore, the optimal solution provided by GIDP for W-subsidy problem in (6.17) may not be equivalent to the optimal solution in (6.14).

Figure 6-11. Comparison of the total profit.
6.5 Performance Comparison for Overnight and Daytime Charging Scenarios

This section compares performance of the GIDP and a heuristic greedy algorithm (HGA) for two charging scenarios: overnight (at-home charging) and daytime (parking-lot charging). Aggregator profit, delayed charging, and battery degradation are evaluated. In addition, a sensitivity analysis is conducted in the overnight charging scenario to assess the impact of the ESS capital cost and the average frequency regulation service price on aggregator profit.

The differences between the overnight case (at-home charging) and the daytime case (parking-lot charging) are: 1) the parameter setting of the penalty cost function for delayed charging are different. In the overnight charging case, the penalty cost for delayed charging is less because the probability of a PEV leaving the garage in the middle of the dispatching period is low. For example, for an at-home charging PEV with a dispatching period between 8:00 p.m. and 6:00 a.m., it is very unlikely that the vehicle will leave the garage before 5:00 a.m. As a result, delayed charging poses little risk on unfulfilling a customer’s charging requests. However, in the daytime charging case, where the dispatching period is between 8:00 a.m. and 5:00 p.m., the vehicle owner has a much higher chance to drive the vehicle for attending unscheduled events. Thus, delayed charging is more likely to cause unfulfilled charging request at the time of departure. 2) PEVs departure probability are different. In the overnight charging case, the departure probability of a PEV in the dispatching period [0,T] is very low so I assume no PEV will depart during [0,T]. In the daytime parking lot charging scenario, the parking time of a PEV follows a uniform distribution between [0.5h,5h],
representing a high chance that the PEV will depart from the parking lot during the dispatching period.

6.5.1 Case Setup

A mixed group of 1000 PEVs (e.g., vehicles with and without V2G functions) are considered in the case studies so that the benefit of having V2G functions can be studied. The capacities of the Li-ion batteries are uniformly distributed between 10 and 50kWh. The rated power levels of the PEV chargers of the mixed group are listed in Table 6-4. Those values are defined by the Society of Automotive Engineers (SAE) [68].

Table 6-4. Charger rated power and mix.

<table>
<thead>
<tr>
<th>Charging Ratings</th>
<th>Rated Power</th>
<th>Mix</th>
</tr>
</thead>
<tbody>
<tr>
<td>AC level 1</td>
<td>1.4 kW (12 A), 1.9 kW (20 A)</td>
<td>35%, 35%</td>
</tr>
<tr>
<td>AC level 2</td>
<td>4 kW (12 A), 19.2 kW (12 A)</td>
<td>25%, 5%</td>
</tr>
</tbody>
</table>

The scenario settings are described in Table 6-5. Note that in both case studies, I considered four V2G penetrations from 40% to 100%.

Table 6-5. Scenarios of charging control.

<table>
<thead>
<tr>
<th>Scenarios</th>
<th>Characteristics</th>
<th>Percent of V2G* vehicles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overnight (at-home) Charging Period 11 p.m. to 5 a.m.</td>
<td>No departure action.</td>
<td>40%, 60%, 80%, 100%</td>
</tr>
<tr>
<td>Daytime (parking-lot) Charging Period 12 p.m. to 5 p.m.</td>
<td>All PEVs come and depart the parking lot randomly over the whole service time</td>
<td>40%, 60%, 80%, 100%</td>
</tr>
</tbody>
</table>

*The V2G vehicles are evenly distributed on the four power levels in Table 6-4.
The 4 MW regulation up signal (See Figure 6-12) obtained from California Independent System Operator (CAISO) is used for comparison and analysis. The frequency regulation service price is assumed to be $35/MW in the first two scenarios.

![Figure 6-12. 4MW Regulation signal in California ISO used for comparison and analysis.](image)

In the first two scenarios, I used $500 /kWh as the unit price of the Li-ion battery. This assumption is made because although the unit price of the Li-ion battery ranges from $1000 to $1450/kWh [81, 82] now, the U.S. Advanced Battery Consortium predicts that the cost of battery may drop to $200-$300/kWh by 2016 [81].

### 6.5.2 Heuristic Greedy Algorithm

The HGA follows two basic rules. The first rule is that the priority list of PEV for participating in regulation service is determined by the SOC of the PEV battery. Batteries with higher SOCs have higher priorities. This is because battery degradation cost decreases monotonically when SOC state increases. The second rule is that idling has a higher priority over discharging. This is because idling does not incur battery degradation cost. The dispatch logic of the proposed HGA is presented in Table 6-6.

98
Table 6-6. Dispatch logic of the heuristic greedy algorithm.

1. The default command to be sent to all PEVs at the beginning of a decision step $j$ ($j \in [1, n_r]$) is “charge”.

2. At the beginning of each step, sort all PEVs in descending order based on their current SOC to obtain the priority list.

3. If $\sum_{i=1}^{N_{PEV}} (p_i / \eta_i) \geq r^{UP}(j)$,
   
   3.1. Switch PEV $i$ from charging to idling.
   
   3.2. $r^{UP}(j) = r^{UP}(j) - p_i / \eta_i$.
     
     If $r^{UP}(j) \geq 0$, $i=i+1$, and go back to step 3.1.
     
     Else Stop.

4. If $\sum_{i=1}^{N_{PEV}} (p_i / \eta_i) < r^{UP}(j)$,
   
   4.1. Switch PEV $i$ from charging to discharging.
   
   4.2. $r^{UP}(j) = r^{UP}(j) - p_i / \eta_i - p_i / \eta^d$.
     
     If $r^{UP}(j) \geq 0$, $i=i+1$, and go back to step 4.1.
     
     Else idling all PEV $i+1$ to $N_{PEV}$ and stop.

6.5.3 Case Study Results

The aggregator profits calculated by GIDP and HGA for the overnight and daytime charging cases are shown in Figure 6-13 and Figure 6-14.
Figure 6-13. Aggregator profit calculated by GIDP and HGA for the overnight charging case.

Figure 6-14. Comparison of aggregator profit from GIDP and HGA in daytime charging case.
In each scenario, four V2G penetration levels is compared, as defined in Table 6-4. The following observations are made:

- As shown in Figure 6-13, in the first couple of hours, the aggregator profits are similar for all cases. This is because at lower SOCs, “idling” costs less than “discharging”. Thus, the GIDP priority list is similar to the HGA priority list. However, as the GIDP considers not only the SOC level of the PEV at step \( j \) but also the unmet SOC \( (s'_i - s_i(j)) \), the compensation to a PEV for charging delays can be less. Therefore, the overall profit gain of the GIDP over the HGA increases consistently over the whole scheduling period.

- In the overnight charging cases, the aggregator profits increase with respect to service time. This is because when the SOC of the PEV battery increases, the cost of charging delays and battery lifetime degradation decreases. Thus, the aggregator will receive more profit in the last couple of hours.

- The number of V2G vehicles in both cases does not significantly affect the aggregator profit. This is because the idling selections have a higher priority over discharging selection in HGA and GIDP only discharge vehicles when it has relatively lower degradation cost at relatively higher SOC.

Figure 6-15 and Figure 6-16 show the accumulated battery degradation cost comparison between GIDP and HGA in the overnight and daytime charging scenarios.
Figure 6-15. Comparison of battery degradation cost from GIDP and HGA in scenario 1.

Figure 6-16. Comparison of battery degradation cost from GIDP and HGA in scenario 2.

From the results, we have the following observations:
• Because the degradation cost only occurs when the vehicle battery discharges, the degradation cost in the HGA case represents the minimum degradation cost for participating the regulation service (HGA prefers “idle” over “discharge”).

• GIDP schedules “discharge” over “idle” when the cost of discharging a PEV is lower than letting it idle (See Figure 6-8), the degradation cost will be higher than the HGA but the overall payment from the aggregator to the PEV is lower because it costs more to pay for the delayed charging.

• The degradation cost is flattened after 4 hours in overnight charging cases. This is because the vehicles have very low probability of departing from the garage, so most of them will be charged to higher SOCs after 4 hours, making the degradation cost negligibly low. In a daytime charging case, because a vehicle can depart the garage randomly over the charging period, the probability of discharging a battery at a lower SOC is higher than in the overnight charging case. As a result, the degradation cost of the daytime charging case is higher than that of the overnight charging case. In addition, the cost will not be flat at the end of the charging period.

• When there are more V2G vehicles in the fleet, the degradation cost will increase when using GIDP and decrease when using HGA. This is because the GIDP turns to schedule “discharge” over “idle” when relatively higher SOCs are reached.

To compare the payment to V2G vehicles and non-V2G vehicles, the 50% V2G vehicles in the overnight charging scenario is selected as examples. As shown in Figure 6-17, the payment to the V2G users is generally higher because the V2G users’ can provide twice as much service so the service rewards are higher. In addition, there is a battery degradation cost. Initially, when the SOCs of most of the cars are low, most V2G vehicles provide the
frequency regulation services by idling. So they receive the same payments as the non-V2G vehicles. After a few hours, there are more and more V2G vehicles providing service by discharging, so the payment to V2G vehicles will increase. Compared with HGA, GIDP results in less payment for delayed charging compensation because GIDP allows the PEVs with lower SOCs to continuously charge without interruption by discharging V2G vehicles at high SOCs.

![Comparison of payment to V2G vehicles and non-V2G vehicles from GIDP and HGA in scenario 2.](image)

Figure 6-17. Comparison of payment to V2G vehicles and non-V2G vehicles from GIDP and HGA in scenario 2.

To summarize, HGA is a conservative dispatch method that treats “discharge” as the last resort to avoid the degradation cost. GIDP is an aggressive dispatch method that allows V2G vehicles to discharge at higher SOCs when the overall revenue outweighs the degradation cost.

Thus, for a given regulation signal at a given price, the aggregator’s profit depends on two key factors: how much payment they should make to a customer for delayed charging and how much they should make to a customer for degradation cost. Delayed charging is a measure of comfort and selecting a group of customers with low expectations for
compensations for unfulfilled charging requests and a high probability to stay at a charging station for the entire scheduling period is the key to lowering such a comfort payment. Degradation is determined by the laws of physics and its cost depends on battery costs and the operation and maintenance costs.

6.5.4 Sensitivity Study

Assumptions made for the regulation service price and battery cost can affect aggregator profit [81, 82]. Therefore, I did a sensitivity study on the aggregator’s profit with respect to varying regulation service prices using the 8-hour overnight scenario as an example. The aggregator profit at regulation price $10~$50/MW is compared for a case where Li-ion battery unit price is $500/kWh and the same 4-MW regulation signal shown in Fig. 6 is used, and the results are shown in Figure 6-18. As expected, for both algorithms, the aggregator profit increases linearly with respect to the regulation service price.

Aggregator profit with respect to battery unit price when using the GIDP method at an regulation service price of $35 MW is plotted in Figure 6-19. When the unit price increases from $100/kWh to $500/kWh, the aggregator’s profit decreases rapidly. But after $500/kWh, the curve is flattened. This is because the cost of degradation becomes so high that discharging the vehicle battery is no longer profitable. Thus, GIDP will dispatch all V2G PEVs to idling rather than discharging and the impact of the battery unit price will saturated and will have no further influence on the aggregator’s profit.
Figure 6-18. Profit of GIDP and HGA under different regulation service prices.

Figure 6-19. Profit of GIDP and HGA under different battery unit costs.

6.6 Conclusions

In this chapter, I present a real-time, greedy-index based dispatching policy (GIDP) for dispatching PEVs to maximize the aggregator’s profit by providing frequency regulation service. A new benefit mechanism is proposed for PEVs based on the amount of regulation service they provide, compensation for delayed-charging, and reduction of battery lifetime because of battery degradation. The optimal PEVs dispatch problem is extremely hard to solve because of the huge amount of state variables caused by high dimensional space and
uncertainty in PEVs arrival/departure times, as well as the unpredictable frequency regulation signals. The proposed GIDP simplifies the solving process by transforming the optimization problem from high-dimensional space with uncertainty to a new one-dimensional space while preserving the optimum of the original problem. Simulation results show that the GIDP can be used to provide profitable and robust frequency regulation service with minimal impact on PEVs’ convenience and battery lifetime.
CHAPTER 7 Conclusions and Future Works

This thesis presents my research efforts on developing and advancing the solution methods for solving large-Scale power system resource scheduling and dispatch problems. The work was sponsored by PNNL as part of the effort of developing an production cost model EOM for studying the renewable integration problems, addressing climate change impacts on power grid operations, co-optimization of hydro-thermal operations, and dispatching distributed energy resources. The developed methodologies have been integrated into the EOM tool and summarized in Figure 7-1. The first part of my effort focuses on enhancing the MILP-UC solving speed so that it is possible to run a great number of case studies in order to address the uncertainties in renewable resources, a wide range of weather patterns, as well as increased number of operational constraints at finer time resolutions. The second part of my effort focuses on extending the EOM modeling capability to include dispatching distributed energy resources for providing ancillary services.

Figure 7-1. Modeling framework of EOM
In Chapter 3, a new MILP formulation has been proposed to solve the UC problem in large-scale power grid. The proposed formulation requires fewer binary variables and constraints than previously models, yielding a significant improvement in computational speed. Based on pattern analysis of UC results and sensitivity of UC results to different costs, in Chapter 4, I presented a modified priority list (MPL) based MILP method to decrease the MILP UC problem scale and solving difficulty by truncating the vast option space of the binary decision variables (always “on” generators and always “off” generators) to a smaller set of variables that can be efficiently solved in a MILP solver. After the problem reformulation and simplification, a case study is presented in Chapter 5 to quantify the impacts of heat waves on power system reliability and production cost. Then, in Chapter 6, I extend the scheduling and dispatch problem from generation resources to distributed energy resources: dispatching plug-in electric vehicles. A real-time, greedy-index based dispatching policy (GIDP) was proposed to dispatch distributed energy resources (providing ancillary services using plug-in electric vehicles (PEV)). The proposed GIDP transforms the PEV dispatching problem from a high dimension space with uncertainty to a new one-dimensional space while preserving its solution optimality. By solving the transformed problem in real-time, the global optimality of the GIDP solution can be guaranteed by the mathematical proved “indexability”.

My future work will be to further expand EOM’s capability by investigating co-optimization algorithms for linking the hydro scheduling process with the power system scheduling process. This will enable the EOM tool to model the hydro-thermal coordination when integrating wind and solar into the power grids. Global warming not only causes average temperature to rise but also leads to more frequent and severe extreme weather
events such as heat waves and droughts. These extreme weather events have significant impacts on water availability for both non-power and power usage, and can greatly impact hydro power generation and thermal power plant cooling. In addition to the decrease in energy generation from hydro power generators, the increasing temperature reduce the thermal power plant cooling efficiency because higher cooling water inlet temperature leads to less efficient cooling and potentially higher outlet temperatures, which are strictly limited by Environmental Protection Agency (EPA) regulation. During the peak load hours, the insufficient water for thermal cooling could cause extra high cooling cost and severe power shortage problems.

The integrated production cost model with high spatial and temporal resolution is required to study the nexus between climate, water and energy system. Existing research [83-85] works mainly focus on the evaluation of water availability due to climate changes and the corresponding impacts on power generation capacity of thermal and hydro plants. These studies are incomplete because they do not include the impacts of climate, water system on energy system, which are actively respond to the changes in temperature and water resources.

Therefore, my future research effort will be to develop a hydro-thermal co-optimization algorithm that interacts with a water management tool so that the monthly available hydro energy provided to the EOM by the water management tool can be used to optimize the hydro generation schedule under its minimum and maximum generation constraints. Then the hydro generation scheduling can be feed back to the hydro model. Through such an iterative process, the global optimal result can be reached for both the water management system and power system scheduling problems.
REFERENCES


