Spring 5-18-2019

# Operation and Planning of Data Centers in Electricity Networks

Ali Vafamehr
*Southern Methodist University*, avafamehr@smu.edu

OPERATION AND PLANNING OF DATA CENTERS IN ELECTRICITY NETWORKS

Approved by:

Dr. Mohammad Khodayar

Dr. Jianhui Wang

Dr. Behrouz Peikari

Dr. Khaled Abdelghany

Dr. Eli Olinick

# OPERATION AND PLANNING OF DATA CNTERS IN ELECTRICITY NETWORKS

A Dissertation Presented to the Graduate Faculty of the

Bobby B. Lyle School of Engineering

Southern Methodist University

In

Partial Fulfillment of the Requirements

For the degree of

Doctor of Philosophy

With a

Major in Electrical Engineering

By

Ali Vafamehr

(B.S., University of Tabriz, 2009)
(M.S., Urmia University, 2013)

May18, 2019

TO MY DEAREST PARENTS AND BROTHER FOR THEIR LIFELONG LOVE AND ENCOURAGEMENT

AND

TO MY LOVELY GIRLFRIEND FOR HER UNCONDITIONAL LOVE AND SUPPORT

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my advisor Professor Mohammad Khodayar for the continuous support of my Ph.D. studies and related research and for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. It would not have been possible to complete and accomplish this thesis without Dr. Khodayar's mentorship. Besides my advisor, I would like to thank the rest of the members of my thesis committee: Prof. Jianhui Wang, Prof. Behrouz Peikari, Prof. Khaled Abdelghany, and Prof. Eli Olinick for their insightful comments and encouragement.

During my time at SMU I was fortunate to make lifelong friends with several talented students. Together we experienced many hours of comradery, lively discussions, and we supported each other during the rigors of graduate school. I sincerely value their friendship and the kind support they gave me during my studies at SMU.

Last but not the least; I would like to thank my family. Firstly, I would like to thank my parents Foruzandeh Haji-Akbari and Ahad Vafamehr and my dear brother Farhang Vafamehr, for supporting me emotionally throughout writing this thesis and in my life in general. Through many Skype phone calls and texts, they were able to support me through every step of my studies, although they are thousands of miles away in Iran. I extend this gratitude to all my family members in Iran, especially my 97-year-old grandmother. Lastly, I would like to thank my girlfriend Martha Penturf and her family: Dr. Mary Hise, Dr. John Brown, and Charlie Penturf. Whether it was Martha listening patiently at the end of a long day, or the family welcoming me as their own at a holiday gathering, they have all been generously supportive.

Vafamehr, Ali B.S., University of Tabriz, 2009
M.S., Urmia University, 2013

Operation and planning of Data Centers in Electricity Networks

Advisor: Mohammad Khodayar

Doctor of Philosophy degree conferred April 26, 2019

Dissertation completed May 18, 2019

First chapter of this dissertation introduces the contents, related literature review and contributions of each chapter briefly.

Second chapter of this dissertation presents a coordinated expansion planning for data centers in the data and electricity networks considering the uncertainties in the planning horizon to ensure an acceptable rate of service to the requests received from the end-users in the data network. The proposed problem addressed the uncertainties in the expansion planning of the electricity networks including the availability of renewable generation resources, the variations in electricity demand, the availability of generation and transmission components in the electricity network, and the uncertainties in the number of requests received by the user groups in the data network.

The objective is to determine the location and capacity of the data centers as well as the required data routes while considering the imposed constraints in the electricity and data networks. The installation cost of data centers and data routes, as well as the expected operation cost of the data centers, are minimized. The problem is formulated as a mixed integer linear programming problem and Bender decomposition and electricity price signals are used to capture the interaction among the data and electricity networks.

Third chapter proposes a framework for the expansion planning of battery energy storage in distribution electricity network with the data center facilities. The objective of the distribution system operator is to minimize the installation costs of battery energy storage as well as the operation cost of the distribution network. Data centers are operated by the data center operators to provide cloud-computing

services. Data center operators leverage the flexibility in the data center demand to minimize the energy costs. The coordination between data center operators and the distribution network operator would minimize the expansion planning costs of the battery energy storage facilities in the distribution network. The proposed framework leverages Benders decomposition technique to address the interactions between the data center and distribution network operators.

Fourth chapter highlights the interdependence among the electricity and cloud computing infrastructure systems by emphasizing energy-aware cloud computing. Cloud computing, as a trending model for the information technology, provides unique features and opportunities including scalability, broad accessibility and dynamic provision of computing resources with limited capital investments.

This Chapter presents the criteria, assets, and models for energy-aware cloud computing practices. Energy management practices for cloud providers at the macro and micro levels are introduced that improve the cost and reliability of cloud services. Moreover, fourth chapter envisions a market structure that addresses the impact of the quality and price of energy supply on the quality and cost of cloud computing services.

Cloud computing as an emerging computing model provides computing resources as general utilities for the end-users through the Internet. Fifth chapter presents a market structure to address the competition among cloud providers in the wholesale electricity market and cloud computing market (CCM). The cloud providers compete to serve the customers in the CCM while they bid for the electricity demand in the wholesale market. The proposed market structure highlights the interdependence among electricity and cloud computing infrastructure systems by introducing the dynamic game with complete information among cloud providers in the wholesale market and the CCM. At each infrastructure system, the operation strategies of the market participants are determined by forming bi-level optimization problems in which the upper-level problem maximizes the payoff of the market participants while the lower-level problem represents the corresponding market settlement process. Independent system operator ensures the security and reliability of the network in the wholesale market while the demand and supply are balanced in the CCM to determine the price of the offered computing resources.

In sixth chapter, a stochastic operation framework is proposed to minimize the operation cost of the distribution networks by leveraging the demand flexibility of data centers. The emerging distribution networks are equipped with renewable generation resources, controllable loads and corresponding monitoring and control assets to regulate the aggregated demand actively in the bulk power system. The volume of the energy received from the bulk power system as well as the price of electricity contributes to the operation cost of the distribution network. Distribution system operators balance the demand and supply and compensate for the intermittency and variability of the local renewable generation resources and non-controllable loads using the controllable generation and demand assets. Data centers are large flexible electrical loads in distribution networks that could help the distribution system operators to manage the operation cost and to mitigate the imbalance between the demand and generation. The introduced uncertainties in the operation horizon provide economic risks in the distribution network operation. Hence, providing efficient measures for risk associated with the uncertainties provides an insight for the decision makers to avoid over conservative operation decisions. The proposed risk averse operation-planning framework constrains the volatility of the expected cost through the conditional value at risk (CVaR) assessment.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

Chapter 1

# INTRODUCTION

## 1.1 A Framework for Expansion Planning of Data Centers in Electricity and Data Networks under Uncertainty

Internet Data Centers (IDCs) are the physical layouts of the computing clouds that are equipped with thousands of devices including switches, routers and several types of servers to provide various services to the end-users [2.1]. Considering the massive scale of IDCs, a considerable amount of energy is consumed by these entities that needs to be considered in the electricity network operation planning. In 2011, the energy consumption of IDCs was approximately 1.5% of the total electricity consumption worldwide and was increased by 56% from 2005-2010 [2.2]. The challenges associated with the reliability and economics of energy supply for the IDCs are further highlighted with the increasing proliferation of cloud computing. Energy-related costs are estimated as 46% of the operation cost of an IDC [2.3]. Energy management solutions such as energy efficiency in chip multiprocessing [2.4], network power management [2.5], online control for power supply system [2.6], and storage power management [2.7] are among the major efforts to conserve energy in the IDCs. Virtual machine (VM) live migration technology [2.8] enables spatial shifting of workloads among servers through multiple VM deployments. Consequently, IDCs are envisioned as large and flexible electrical loads that facilitate demand response practices to reduce the total peak demand by up to 20% [2.9]–[ 2.14]. The study in [2.15] investigates the potential benefits of flexible IDC power management utilizing local fuel cell generation. Other studies investigated the benefit of spatial shifting of the data centers' workloads to the locations with cheaper energy or abundant renewable resources [2.16] and [2.17]. There are generally two categories of workloads: delay-intolerant and delay-tolerant workloads. Delay-intolerant workloads such as web services have limited flexibility in temporal shifting whereas the delay-tolerant workloads such as CPU-intense batch computing jobs are shiftable to the periods in which the energy is cheaper [2.18]–[ 2.20]. Furthermore, load shedding considering the required Quality-of-Service

(QoS) can reduce the energy consumption and associated costs [2.21]. The research work reported in [2.22]–[2.25] examined the coordinated operation of data centers and renewable resources to minimize the carbon footprint and the operation cost. The expansion planning of data centers in electricity and data networks was addressed in [2.26]. Here, the objective was to find a deterministic solution for the expansion planning while the operation cost of the data centers, as well as the uncertainty in the planning horizon were ignored. While the earlier research is focused on developing approaches to address the operation schedule of the IDCs in the electricity market, there are limited research efforts dedicated to the expansion planning of the IDCs in the electricity and data networks. The IDC expansion planning reported in [2.27] addresses the capacity allocation of the future site constructions/expansion of data centers to meet the demand in the data network. Such practice is important as the future data centers are expected to be established in areas with lower electricity prices, least cost for data routes and bandwidth capacity, and lower carbon footprint and environmental effects. The outdated data centers, insufficient capacity of current data centers, the adventure of big data, virtualization and new applications outsourced to cloud services highlight the merit of the expansion planning for the IDCs [2.28]. Effective expansion planning strategies are further underlined as the traditional perceptions of IDC facilities – that were low-density and site-constructed, with considerable planning lead time – were transformed toward flexible, more rapidly deployable and modular assets. Expanding the capacity of IDC as new demand entity in the electricity network that operates close to its capacity limits, will reduce the reliability and security of energy supply [2.29]. Hence, the capacity expansion strategies for IDCs should capture the reliability and security of energy supply as the quality of service provided by the IDCs to the cloud users, is affected by the quality of utilized energy.

## 1.2 Network Constrained Expansion Planning of Battery Energy Storages in Distribution Network with Data Centers

The growing demand for cloud computing and large investment of companies (e.g. Google, Microsoft, Facebook) in data centers to support cloud services will increase the electricity demand in the distribution networks. Data centers are composed of Information Technology (IT) equipment including servers, storage devices, network switches and routers that form the physical layer of the cloud computing architecture. In 2014, these assets consumed 70 billion kWh of energy which is approximately 1.8% of the

total energy consumption in the United States [3.1]. The electricity demand of data centers has increased by 4% from 2010-2014 and it is expected that the rate of increase remains the same for 2014-2020 [3.1].

Despite the considerable energy consumption, data centers offer a unique flexibility in how and when power is consumed. Geographically distributed data centers facilitate shifting the workloads spatially. On the other hand, the process of delay tolerant workloads allows temporal shifting while honoring customers' service level agreements. Such flexibilities enable the operators to implement various energy management solutions with different objectives, such as reducing the energy costs and carbon footprint, and improving the sustainability of cloud computing services [3.2]- [ 3.4]. Furthermore, to reduce the adverse environmental impacts of the data centers, the cloud providers are taking various initiatives to leverage renewable energy resources to promote green cloud computing [3.5].

Among renewable energy resources, solar photovoltaic (PV) generation is gaining more attention in distribution networks as the installation cost is decreasing due to technological advancements and subsidiary energy policies. The cost of solar modules has decreased by 80% since 2009, and therefore, the cost of energy produced by these modules has decreased by 75% from 2010 to 2017 [3.6]. While coupling these sustainable energy resources with data centers could improve the emission and cost measures, the intermittency and variability of these energy resources as well as the uncertainty in demand of data networks introduces operational challenges for the Data Center Operators (DCOs). DCOs operate one or more data centers in a distribution network and ensure the reliability and security of cloud computing services while minimizing the cost and/or adverse environmental impacts of the data centers [3.7].

Furthermore, the integration of renewable resources introduces operational challenges to the distribution network operators (DNOs) to maintain the stability, reliability and power quality [3.8]. The considerable demand of data centers along with their uncertainty and variability will further intensify these operational challenges. The DSO ensures the reliability and security of the distribution network by capturing the uncertainty in generation and demand and the availability of the distribution network assets into its operation planning framework. The objective of the DSO includes minimization of operating cost, network loss, and/or voltage deviation to ensure the efficient energy consumption and secure operation of the distribution network.

The peak demand of the data centers has considerable effect on the capacity expansion of distribution assets. The increase in the peak demand will increase the capital cost of expansion plans as the size of the power delivery equipment including transformers and distribution lines are dependent on the peak power consumption of the data centers. Furthermore, considerable energy consumption of data centers will increase the operating cost of the distribution network and reduce its power delivery capacity margin. It will also reduce the acceptable voltage margin at the distribution network buses. The DSO leverages flexible generation resources including distributed energy resources (DERs), battery energy storage (BES), flywheels and capacitors to mitigate the adverse effects of data centers' demand as well as the variability and uncertainty in renewable generation assets.

The short response time of the BES systems makes them valuable assets to improve the distribution network operation measures. BES could be used for energy arbitrage, increasing the capacity of renewable energy resources, deferring investment in the distribution network components by relieving the network congestions, reducing the carbon footprint, reducing the loss in the electricity network, and facilitating the transient and steady state voltage control. High investment costs of BES impede the wide-spread allocation of these resources in the distribution networks and an effective expansion planning strategy for sizing and siting of BES in the distribution network is crucial to maximize the benefits of this technology in the distribution networks.

Earlier research in this domain is focused on the expansion planning of BES in distribution networks. The expansion planning of BES for energy arbitrage is addressed in [3.9], [3.10] where the BES systems store electricity in low-price periods and release the energy in high-price periods. Expansion planning of BES for congestion management in distribution feeders is proposed in [3.11]-[ 3.13]. By reducing the peak demand, BES reduce the power flow in distribution lines and impede congestion in distribution feeders. Such strategies could further defer the investments in distribution assets. The expansion planning of energy storage systems to improve the reliability indices was addressed in [3.14]- [ 3.16]. In [3.14]- [ 3.15], the expected energy not supplied was captured while in [3.16] the momentary average interruption frequency index was considered as the reliability index. The expansion planning of the energy storage systems to improve the voltage regulation is addressed in [3.15] and [3.17]. Here, the injected real and reactive power

of energy storage system enables the DSOs to regulate the bus voltages to prevent voltage instability. In contrast to reliability indices, [3.18] and [3.19] consider expansion planning of energy storage systems to reduce the loss in the distribution network. This is achieved by reducing the peak demand as well as the power flow in the distribution lines and transformers that eventually reduce the energy loss in the distribution network. In [3.20]- [3.22] the objective of the expansion planning for energy storage system is to improve the dispatchability of the renewable resources by absorbing their surplus energy in low demand periods and releasing it at high demand periods. In [3.22], the joint expansion planning for energy storage systems and distributed generation units considers the emission cost of generation units in the objective function to decrease the carbon footprint in the electricity network.

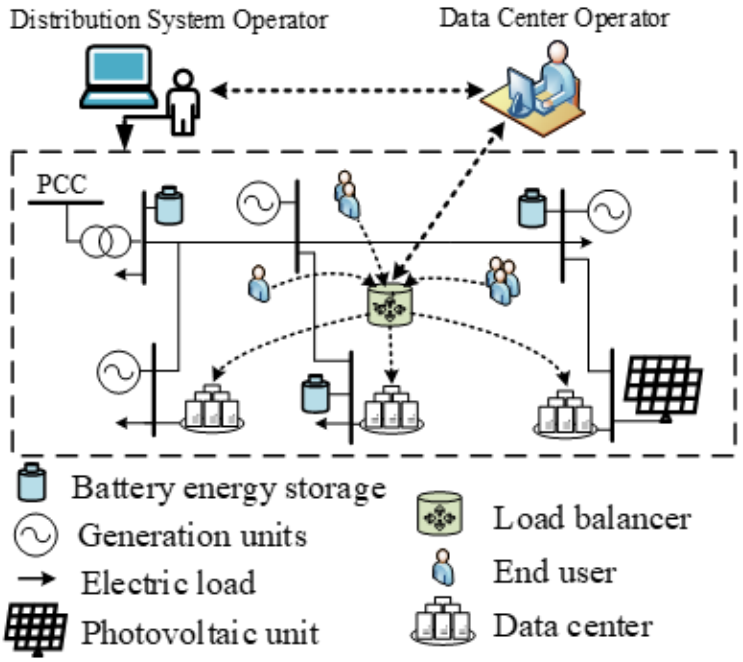


Figure 1-1 Distribution network with data centers

Data centers are flexible electric demands as the power consumption of a data center is determined based on the processed workloads. The temporal shifting of the delay-tolerant workloads from peak hours to off peak hours will reduce the operational cost. Furthermore, shifting the data workload among the data centers improves the operational measures in the distribution network and minimize the mismatch between the demand and supply considering the imposed uncertainties in the distribution network. Figure 1-1 shows

the physical structure of the distribution network with the data centers. As shown in this figure, the DCO regulates the workload among the data centers to minimize the operational cost. The DCO collects the requests of the end users in the data network and distributes the requests among the data centers, considering the data centers' computing resources and the network bandwidth constraints. The allocated requests are sent by routers and network switches to the data centers where these requests are processed by servers.

The proposed expansion planning for energy storages in [3.9], [3.10], [3.12]-[ 3.14], [3.16]-[ 3.18], [3.20] and [3.22] leverage heuristic methods including genetic algorithms and particle swarm optimizations to solve the planning problem. Heuristic methods are based on random and iterative procedures, which may not guarantee the optimal solution. The problem formulation in [3.21] is MINLP, which is solved by CONOPT solver so the optimal solution heavily depends on the type of nonlinearity. In [3.15], a multistage MILP expansion planning for energy storages, conventional distributed generations (DGs) and renewable resources is proposed to maximize renewable integration while the power quality and stability is maintained within the required levels. In [3.15] the uncertainty in renewable generation and demand is addressed by assuming two scenarios with ±5% deviations for predicted values.

The studies in [3.11] and [3.19] leveraged simplified AC power flow formulation. The problem formulation in [3.11] and [3.19] solves a linear objective function with second order cone constraints for the power flow to improve the accuracy of the solution. This formulation poses significant computational burden when applied to stochastic settings.

The expansion planning in [3.9]-[3.14], and [3.16]-[3.22] capture the sizing and siting of energy storage while ignoring the installation time, and other works in [3.9]-[3.11], [3.13], [3.16], [3.17] and [3.19]-[3.21] are focused on deterministic solution for the expansion planning problem ignoring the uncertainty in the planning horizon. The uncertainty in distribution network demand is addressed in [3.12] and [3.18], while the uncertainty in the characteristics of the batteries including the technical and economic characteristics of the battery technologies are considered in [3.14].

## 1.3 Energy-Aware Cloud Computing

The emerging cloud-computing model facilitates access to computing resources for end users through the internet. Cloud computing is a model that enables on-demand access to the shared pool of customizable computing resources (e.g. servers, storage, networks, and applications) and services [4.1]. These resources can be rapidly deployed with minimal management efforts and marginal interactions with the service providers. Providing dynamic computing resources in the cloud computing paradigm enables corporations to scale up/down the provided services, considering their clients' demand and the cost of the leveraged resources that contribute to the operational cost of the information technology (IT) facilities. The scalability of the cloud services enables smaller businesses to benefit from different categories of expensive computing-intensive services that were once exclusively available to large enterprises. Cloud computing remedies the IT barriers, especially for small and medium-sized enterprises, and provides efficient and economical IT solutions as the cloud providers develop tools and skills to exclusively focus on handling the computational and IT challenges. With marvelous effects of cloud computing on the IT industry, large enterprises such as Google, Amazon, and Microsoft endeavor to establish more powerful, reliable, and economically efficient cloud computing platforms. The backbone of cloud computing is data centers. Cloud computing is achieved by establishing distributed data centers that consume a significant volume of energy. Data centers leverage advanced energy management solutions to achieve the targeted computing reliability and economic efficiency.

This chapter presents the envisioned market structure for energy-aware cloud computing that incorporates energy management strategies at multiple physical layers.

## 1.4 Oligopolistic Competition among Cloud Providers in Electricity and Data Networks

The advent of cloud computing increased the electricity demand of the data centers that are operated by the cloud providers in the power networks. Data centers enable temporal and spatial load shifting and curtailment in the electricity and data networks [5.1]–[5.6]. In the wholesale market, the generation companies (GENCOs) maximize their payoff by strategically bidding for the energy supply while the load service entities, load aggregators, and large customers including cloud providers bid for the consumed energy. The

price of electricity is procured by capturing the demand and supply bidding curves and ensuring the demand and supply balance. In a large-scale Internet data center with condensed server clusters, energy costs are approximately 41.6% of the total operation cost [5.7]. Therefore, the energy management practices in data centers largely impact the operation cost and revenue of the cloud providers. While earlier researches [5.8]–[5.12] addressed the energy management in the data centers, the impact of such strategies on the price of electricity and the competition among the cloud providers in the wholesale market were not addressed. As bulk demand entities, the cloud providers bid in the wholesale market to reduce their electricity costs while they bid in the CCM as service providers to maximize their revenue. In the CCM, the cloud providers offer various pricing and contracting structures for multiple types of cloud services including infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). For example, Amazon computing services are billed on an hourly basis, however, other Amazon services such as Simple Queue Service (SQS) and data storage are billed based on the input/output rate of data transfer [5.13], [5.14]. Google pricing schemes are determined by application on a monthly basis [5.15]. The competition among multiple cloud providers to provide IaaS is addressed by an analytical study on the monopoly, duopoly, and oligopoly competition in [5.16]. In [5.17] a three-tier model for the cloud computing is proposed to investigate the interaction between the end users and the cloud providers. Several studies including [5.18] and [5.19] addressed the strategic behavior of the cloud providers and pricing schemes in the CCMs while ignoring the electricity network. SHARP [5.20], Tycoon [5.21], Bellagio [5.22], and Shirako [5.23] are some of the research projects that proposed market structures for trading resources on networked infrastructure systems such as PlanetLab [5.24]. A global cloud computing market structure is proposed in [5.25] to mitigate the challenges that restrict the customers to switch to other cloud providers including inflexible pricing mechanism of the cloud providers, inaccessibility to multiple cloud providers for the same service, and incompatible interfaces and protocols to acquire services from the cloud providers. In the proposed global cloud market structure, the brokers are intermediates between the consumers and providers that lease the resources from the cloud providers and sub-lease them to the end-users. The cloud market is cleared in the market directory to maximize the well-being of all participants (i.e., social welfare maximization). Considering the considerable energy consumption and flexibility of the data center electricity demand, data center demand with associated bidding strategy of the cloud providers would impact the locational marginal

price (LMP) of electricity in the wholesale market. As bulk demand entities, the cloud providers bid in the wholesale market to reduce their electricity costs while they bid in the CCM as service providers to maximize their revenue. Therefore, by adjusting the bidding strategy of the cloud providers, their payoff could be maximized in both the CCM and the wholesale market.

## 1.5 Operation of Distribution Networks with Volatile Supply and Controllable Data Center Demand

The operation scheme of the distribution network is transformed from a passive management where the energy is solely provided from the wholesale market to active operation scheme that leverages the local renewable generation, distributed generation, storage, and controllable loads to improve the economics and security measures of the distribution networks.

Renewable resources introduce variability and intermittency at the generation sector in the distribution network. This issue could further be propagated to the bulk power system and wholesale market by introducing fluctuation in the aggregated demand. The considerable penetration level of renewable resources at the distribution network increases the uncertainty in the aggregated load in the bulk power network, which further adds to the complexity of the cost-effective operation solutions in the bulk power network [6.1], [6.2]. In distribution networks, the distribution system operator (DSO) is responsible for maintaining the security of the network by matching the demand and supply and importing electricity from the bulk power system. The load service entities that represent aggregated demands in the wholesale market ensure the adequacy of energy supply for the distribution networks.

The variability and uncertainty in demand imposed by the distribution networks contribute to the uncertainty in the aggregated demand in the wholesale market. Deploying energy storage technologies and demand response practices were proposed to mitigate the volatility and uncertainty in the aggregated demand of the distribution networks [6.3], [6.4].

Data centers are large and flexible electrical loads. The flexibility in data centers' demand can be leveraged to minimize the operation cost of the distribution network. Furthermore, data centers consume significant energy and their power consumption grows 10-12% per year [6.5]. The growth in the data center demand

makes them an important tool for balancing the demand and supply as their demand capacity increases with the growth in the installed capacity of the renewable resources [6.6]. The flexibility in data centers demand is provided by managing the workloads in the data network. Depending on the type of services provided by the data centers, the workloads are categorized into delay-intolerant and delay-tolerant workloads. Delay-intolerant workloads require processing the end-user requests in limited period, examples of such workloads include web searches and web services. For delay-tolerant workloads, the process time is not critical for the end users as long as the workloads are processed before a designated deadline.

In this chapter, a stochastic operation framework is proposed to minimize the operation cost of the distribution networks by leveraging the demand flexibility of data centers.

Chapter 2

# A FRAMEWORK FOR EXPANSION PLANNING OF DATA CENTERS IN ELECTRICITY AND DATA NETWORKS UNDER UNCERTAINTY

This chapter presents a coordinated expansion planning for the IDCs and data routes considering the operation cost of the IDC and the security of energy supply in the electricity network. This chapter is focused on the economic aspects of the expansion planning of IDCs in the data and electricity networks while ensuring the energy supply security. The proposed expansion strategies determine the location and capacity of the IDCs and the required data routes among the User Groups (UGs) and IDCs while ensuring the security of energy supply. The presented coordination between the electricity and data networks ensures the adequacy and security of energy supply for the IDCs. The unavailability of the generation and transmission assets, as well as transmission congestions in the electricity network, would impact the expansion planning strategies for IDCs in the data network.

The proposed expansion-planning framework captures the uncertainties in the electricity and data networks including the variation in demand, the volatility of renewable energy resources, the outages in generation and transmission components, and the fluctuation in the number of requests received by the UGs. Several scenarios capture the uncertainty in the planning horizon. Figure 2-1 shows the proposed expansion planning framework. As the figure illustrates, the Data Center Investor (DCI) procures the expansion plan for the IDCs and corresponding data routes in the data network. The feasibility of such decisions in the electricity network is evaluated by one or multiple transmission-owning utilities (TOUs). The DCI who manages and invests on the IDCs determines the decisions to install the IDC modules and data routes by solving a mixed integer programming (MIP) problem (master problem) that captures the uncertainty in the data network. Such decisions are passed to the TOUs to check for the feasibility of the network considering the demand of proposed IDCs in multiple scenarios (feasibility sub-problems). If any infeasibility exists in each scenario, infeasibility cuts are generated and sent back to the master problem. In other words, the DCI sends the decisions on installing the IDCs with respective electricity loads to the TOU(s). The TOU(s) may accept the proposed installation or propose different suggestion in case there is a deficiency in the energy supply. In this

chapter, the TOUs send capacity signals to the DCI for revising the proposed expansion plan to satisfy the TOUs' constraints [2.30]. The Benders cuts received from the TOUs are the infeasibility cuts that reflect the feasibility of the expansion decisions made by the DCI. Once the decisions are feasible, the network operation problem in multiple scenarios is solved by the Independent System Operator (ISO) considering the imposed demand by the IDCs. As a result, the locational marginal prices (LMPs) of electricity are passed to the DCI to determine the operation cost of the IDCs in different scenarios and to update the expansion decisions if required.

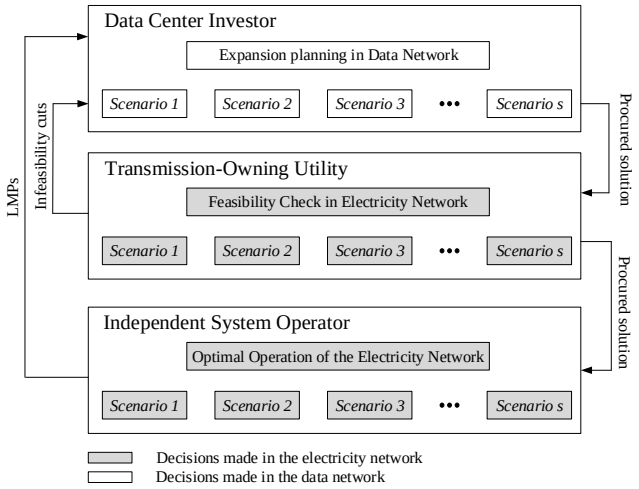


Figure 2-1 The proposed framework for the IDC expansion-planning in electricity and data networks

This iterative process stops once the expansion decisions are feasible and certain stopping criteria for the operation cost are satisfied. It is worth noting that formulating the problem as one optimization problem solved by system operators or DCI is not practically feasible. The system operators in the electricity network (TOU and ISO) neither handle the expansion planning practices for the data centers nor do have any information on the demand of the data network such as the rate of requests received and redirected by the UGs. Furthermore, DCIs have limited access to the information on the electricity network to determine the feasibility of the IDC expansion plans. This chapter proposed a framework to capture the interactions among DCI and system operators (TOU and ISO) for the expansion planning of the data centers in the electricity and data network. The rest of this chapter is organized as follows: Section 2.1 describes the problem

formulation and solution methodology. Sections 2.2 and 2.3 present the case study and summary, respectively.

## 2.1 List of Symbols

**Indices:**

| | |
|---|---|
| $a, b$ | Index of bus |
| $d$ | Index of loads other than data centers |
| $e$ | Type of data route |
| $i$ | Index of data center |
| $j, r$ | Index of user group |
| $l$ | Index of transmission line |
| $n$ | Type of data center module |
| $p$ | Index of periods within a year |
| $s$ | Index of scenario |
| $w$ | Index of wind unit |
| $y, y'$ | Index of year |
| $\phi$ | Index of thermal unit |

**Integer variables:**

| | |
|---|---|
| $m_{i,n}^{p,y,s}$ | Number of active servers in a data center module of type $n$ |

**Binary variables:**

| | |
|---|---|
| $k_{i,n}^{y}$ | Decision variable for installing data center module of type $n$ |
| $h_{i,n}^{y}$ | Auxiliary variable for installing data center module of type $n$ |
| $R_{(.),(.)}^{y,e}$ | Decision variable for installing data route of type $e$ |

**Real variables:**

| | |
|---|---|
| $F_c(.)$ | Generation cost |
| $P_{\phi}^{p,y,s}$ | Generation dispatch of thermal generation unit [MW] |

| | |
|---|---|
| $PL_l^{p,y,s}$ | Power flow of line $l$ [MW] |
| $t_{1,(.)}^{(.)}, t_{2,(.)}^{(.)}$ | Slack variables |
| $\tau_{(.)}^{(.)}, \delta_{(.)}^{(.)}$ | Lagrange multipliers |
| $\theta_b^{p,y,s}$ | Voltage angle of bus |
| $\lambda_{j,i,n}^{p,y,s}$ | Rate of the requests directed by UG j to data center module of type $n$ [request/second] |
| $\lambda'^{p,y,s}_{r,j}$ | Rate of the requests exchanged among the UGs [request/second] |

**Parameters:**

| | |
|---|---|
| $D$ | Desired response time [millisecond] |
| $d_y$ | Annual discount rate |
| $\hat{k}_{(.)}^{(.)}$ | Decision made in the master problem for installing a data center module |
| $L_j^{p,y,s}$ | The rate of requests received by the UG [request/second] |
| $L_{f,b}, L_{t,b}$ | Set of lines starting from/ending at bus $b$ |
| $M_n$ | Maximum number of servers in data center module of type $n$ |
| $\hat{m}_{(.)}^{(.)}$ | Number of active servers determined in the master problem |
| $N_e$ | Capacity of the data route [request/second] of type $e$ |
| $NT$ | Total number of hours in a period |
| $P_{idle}$ , $P_{peak}$ | Power consumption of server in idle and active modes [W] |
| $P_d^{p,y,s}$ | Demand of consumers other than data centers in the electricity network [MW] |
| $P_w^{f,p,y}$ | Forecasted dispatch of wind unit [MW] |
| $P_\phi^{max}$ | Maximum generation dispatch of thermal generation unit [MW] |
| $q_{(.),(.)}$ | Distance [mile] |
| $U_{(.)}^{(.)}$ | Availability of the electricity network components; 1 for being available and 0 otherwise |
| $X_{b,a}$ | Inductance of line between buses $a$ and $b$ |

| | |
|---|---|
| $\beta_n$ | Installation cost of data center module type $n$ [$] |
| $\gamma_e$ | Installation cost of data route type $e$ [$/mile] |
| $\psi_{(.)}^{b}$ | Set of components connected to bus $b$ |
| $\mu$ | Rate of request processed by each server [request/second] |
| $\rho^s$ | Probability of scenario |
| $\hat{\lambda}$ | Decision made in the master problem for the rate of requests directed to the data center |
| $\hat{\delta}_b^{p,y,s}$ | Price of electricity [$/MWh] |

## 2.2 Problem Formulation

The IDC expansion problem provides solutions for increasing the capacity of the IDCs and data routes to serve the customer demand in the data network. Each IDC module consists of several types of servers including web servers, application servers, and database servers. The type of requests determines the characteristics of the workloads and the loading pattern of the corresponding servers [2.31]. For the sake of simplicity in this chapter, all requests are considered to be of the same type and the workload is dependent on the number of requests received. The UG – also known as load balancer among a group of backend servers – aggregate the requests received from the end-users, re-distribute the incoming traffic among backend servers in the IDCs, and send the processed requests back to the end-users [2.27] and [2.32]. UGs are linked together by the data routes for load balancing, sharing the requests and distribute the workloads among the IDCs. In the proposed problem, DCI invests on expanding and allocating the IDC module and data route capacity to ensure the quality of service in the data network. The capacity of the IDC is determined by the number of installed modules on an annual basis [2.33]. The growth in the electricity demand and the number of requests in the data network are captured in the planning horizon. Figure 2-2 shows the structure of the data network in the proposed problem.

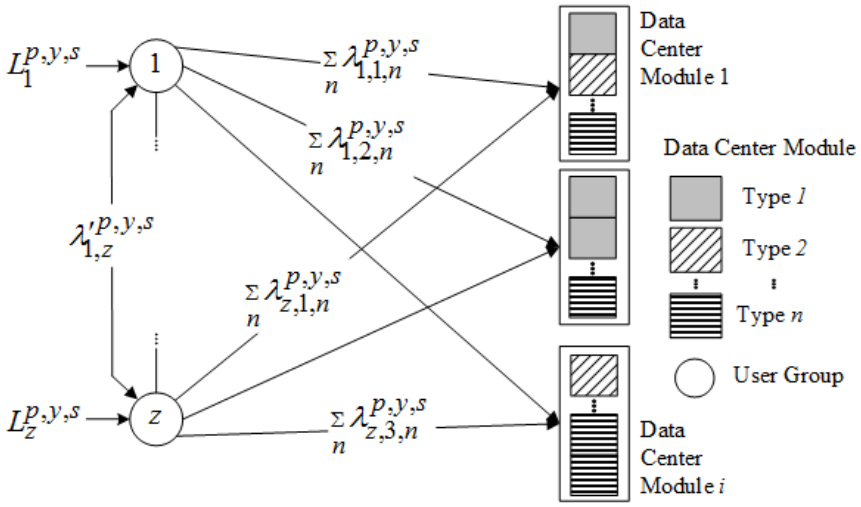


Figure 2-2 Modular structure of data centers with USGs, and data routes.

The figure shows that there are $z$ UGs, each receiving the total input request rate, $L_j^{p,y,s}$ from the end-users in scenario $s$. The UGs communicate with three IDCs. Data center $i$, $i \in \{1, 2, 3\}$ receives requests from the UGj with the rate of $\sum_n \lambda_{j,i,n}^{p,y,s}$ and the rate of exchanged requests among $UG_j$ and $UG_r$ is $\lambda_{j,r}^{p,y,s}$. Data routes leverage fiber optics technology to facilitate fast and reliable communication among the UGs and IDCs. The bandwidth of the data route that connects the UGs to the IDC is determined based on the rate of the directed requests. The number of servers that process the data is dependent on the rate of requests received from the UGs. The queueing model *M/M/1*, with Poisson arrival rate, is used to obtain the average response time. As shown in (1), the average response time captures the average waiting time of requests in the queue and the process time of the IDC. Here, $D$ is the desired response time for the end-users; $P_Q$ is the probability of waiting in the queue, and $P_Q$ = 1once the large volume of data is being processed by the IDCs [2.34].

$$P_Q/(m_{i,n}^{p,y,s}\mu - \lambda_{j,i,n}^{p,y,s}) \quad \leq \ D \tag{2.1}$$

The expansion-planning problem for DCI is an MIP problem that minimizes the expansion and expected operation costs of the IDC and data routes. In this problem, the binary variables are the decisions made to install the IDCs and data routes in the data networks, the integer variables are the number of servers that process the requests in each scenario. Such decisions are made based on the requested data demands by UGs in each scenario. In order to ensure the energy supply security for IDCs in the electricity network, DCI interacts with one or multiple TOUs to ensure the adequacy of energy supply for the installed IDCs. For the

sake of simplicity in this chapter, one TOU interacts with the DCI and Benders decomposition is used to capture such interactions. The TOU checks for the security of the electricity network in each scenario considering the imposed demand for the IDC. If the decisions made by the DCI do not ensure the energy security, the TOU sends the capacity signal in form of Benders cut to change the DCI decisions until the energy supply security is guaranteed. If the decisions made by the DCI ensures the energy security in the electricity network, the ISO solves the economic dispatch to procure the LMPs in each scenario. The LMP of electricity is used by the DCI to assess the expected operation cost of the IDCs and to update the expansion decisions accordingly. In the following sections, the application of Benders decomposition for solving this problem is described and the solution framework to solve the proposed coordinated expansion problem is presented.

### 2.2.1 Application of Benders Decomposition

The general form of the presented expansion planning problem is formulated in (2.2)-(2.6) as an MIP problem where x is the vector of integer and continuous variables and y is the vector of binary variables, respectively. Here, $f(.)$ represents the expected operation cost of the IDCs and $h(.)$ is the installation cost of the IDCs and data routes. The constraints that capture the feasibility of the decisions in the electricity network are shown in (2.4)-(2.5) where, **v** is the continuous decision variable for the electricity network.

$$\min f(\mathbf{x}) + h(\mathbf{y}) \tag{2.2}$$

$$s.t. \quad \mathbf{Ax} + \mathbf{By} \geq \mathbf{c} \tag{2.3}$$

$$\mathbf{Gx} + \mathbf{Hv} = \mathbf{d} \tag{2.4}$$

$$\mathbf{Kv} \leq \mathbf{w} \tag{2.5}$$

$$\mathbf{x} \geq \mathbf{0}, \mathbf{y} \in \{0,1\}, \mathbf{v} \in \Omega_v \tag{2.6}$$

The problem (2.2)-(2.6) is decomposed into a master problem (2.7)-(2.10) and sub-problem (2.11)-(2.14).

$$\min z \tag{2.7}$$

$$s.t. \quad z \geq f(\mathbf{x}) + h(\mathbf{y}) \tag{2.8}$$

$$\mathbf{Ax} + \mathbf{By} \geq \mathbf{c} \tag{2.9}$$

$$\mathbf{x} \geq \mathbf{0}, \mathbf{y} \in \{0, 1\} \tag{2.10}$$

In the master problem (2.7)-(2.10), the upper bound of the objective function is minimized as shown by (2.8). The decision variables **x** and **y** are determined by the DCI once (2.7)-(2.10) is solved [30]. The set of equality and inequality constraints shown in (2.9) captures the physical relationship among the decision variables in the data network. The feasibility of the decision variables in (2.7)-(2.10) is checked in (2.11)-(2.14). Here, the violation in (2.4) as a result of determined decision $\hat{\mathbf{x}}$ in (2.7)-(2.10), is minimized. In order to check the violation of the equality constraint (2.4), the vector of slack variables **t** (i.e., mismatch variables) is minimized. In this formulation, the set of constraints (2.12) and (2.13) captures the physical relationship among the decision variables in the electricity network and these constraints are not shared with the DCI. Therefore, (2.11)-(2.14) is solved by the TOU and if the objective function (2.11) is zero, the proposed DCI solution is feasible in the electricity network. Otherwise, Benders cut (2.15) will be passed to the master problem (2.7)-(2.10). In the next sections, the general formulation that is described above will be formulated for the proposed expansion planning problem.

$$U(\hat{\mathbf{x}}) = \mathbf{1}^T\mathbf{t} \tag{2.11}$$

$$\mathbf{G}\hat{\mathbf{x}} + \mathbf{Hv} + \mathbf{1t} = \mathbf{d} \quad (\boldsymbol{\tau}) \tag{2.12}$$

$$\mathbf{Kv} \leq \mathbf{w} \tag{2.13}$$

$$\mathbf{t} \geq \mathbf{0} \tag{2.14}$$

$$U(\hat{\mathbf{x}}) + (\mathbf{x} - \hat{\mathbf{x}})^{\mathrm{T}}\, \mathbf{G}^{\mathrm{T}}\boldsymbol{\tau} \leq 0 \tag{2.15}$$

### 2.2.2 Master Problem –Expansion Planning in Data Networks

The problem formulation for the coordinated expansion planning of the IDC and data route is shown in (2.16)-(2.25). The vectors of binary, integer and continuous variables are $\mathbf{y} = [\mathbf{k}, \mathbf{R}]$, $\mathbf{x} = [\mathbf{m}, \boldsymbol{\lambda}]$ respectively, where $R_{j,r}^{y,e} \in \mathbf{R}$, $k_i^y \in \mathbf{k}$, $m_{i,n}^{p,y,s} \in \mathbf{m}$, and $\lambda_{j,i,n}^{p,y,s} \in \boldsymbol{\lambda}$. The objective is to minimize the upper bound of the installation cost of IDCs and data routes as well as the expected operation cost of the IDCs, as shown in (2.17). The first term in (2.17) is the installation cost of the IDC modules as well as the capital cost

of the data routes that are installed between the UGs and the IDCs; the second term reflects the installation cost of data routes installed between the UGs and the third term is the expected operation cost of IDCs in which $\hat{\delta}_b^{p,y,s}$ is the LMP of electricity. Here, the rate of requests that is received or processed is associated with the utilized electricity in the IDC and the presented formulation captures a snapshot for each period. As shown in the last term in (2.17), the total energy consumed by servers in each period to process the requests with given rates, is determined by the power consumption of the servers and the length of the period (*NT*). The number of active servers at each period in each scenario, is lower than the installed capacity as shown in (2.18). The number of requests exchanged between the UGs and IDCs in each scenario is limited by the installed capacity of the data routes as shown in (2.19). Similarly, the number of requests exchanged between the UGs in each scenario is limited by the installed capacity of the data routes as shown in (2.20).

$$\min Z \tag{2.16}$$

$$s.t. \quad Z \geq \sum_y \sum_i \begin{Bmatrix} \left((1+d_y)^{1-y}.\sum_n\left(\beta_n.k_{i,n}^y\right)\right) \\ +\left((1+d_y)^{1-y}.\sum_j q_{j,i}\sum_e \gamma_e.R_{j,i}^{y,e}\right) \end{Bmatrix} + \sum_y \left\{\left(1+d_y\right)^{1-y}.\sum_{j,j\neq r}\sum_r \frac{1}{2} q_{j,r}.\sum_e \gamma_e.R_{j,r}^{y,e}\right\}$$

$$+\sum_s \rho^s.\sum_{i\in\psi_i^b} NT.\sum_p\sum_y \left\{\left[\left(1+d_y\right)^{1-y}.\hat{\delta}_b^{p,y,s}\right]\times\begin{bmatrix}\left(\sum_n m_{i,n}^{p,y,s}.[P_{idle}+P_{peak}]\right) \\ +(P_{peak}-P_{idle}).\mu^{-1}.\sum_j\sum_n \lambda_{j,i,n}^{p,y,s}\end{bmatrix}\right\} \tag{2.17}$$

$$m_{i,n}^{p,y,s} \leq \sum_{y'=1}^{y} M_n.k_{i,n}^{y'} \tag{2.18}$$

$$\sum_n \lambda_{j,i,n}^{p,y,s} \leq \sum_{y'=1}^{y}\sum_e N_e.R_{j,i}^{y',e} \tag{2.19}$$

$$\left|\lambda'^{p,y,s}_{r,j}\right| \leq \sum_{y'=1}^{y}\sum_e N_e.R_{j,r}^{y',e} \quad r\neq j \tag{2.20}$$

$$R_{j,r}^{y,e} = R_{r,j}^{y,e} \tag{2.21}$$

$$m_{i,n}^{p,y,s} \geq \left(\mu^{-1}.\sum_j \lambda_{j,i,n}^{p,y,s}\right) + (D.\mu)^{-1}.h_{i,n}^{y} \tag{2.22}$$

$$h_{i,n}^{y} \geq Q^{-1}.\sum_{y'=1}^{y} k_{i,n}^{y'} \tag{2.23}$$

$$L_j^{p,y,s} + \sum_{r,r\neq j} \lambda'^{p,y,s}_{r,j} = \sum_i\sum_n \lambda_{j,i,n}^{p,y,s} \tag{2.24}$$

$$\lambda'^{p,y,s}_{r,j} = -\lambda'^{p,y,s}_{j,r} \quad r\neq j \tag{2.25}$$

Constraint (2.21) ensures the symmetry in the decisions to develop data route among UGs, *j* and *r*. The number of active servers in an IDC, to process the directed requests at each interval in each scenario is determined by (2.22). Here $h_{i,n}^{y}$ is a binary variable that is calculated in (2.23), and represents the installation status of IDC type *n* until year *y*. Since the number of active servers is limited by the capacity of the IDC, $h_{i,n}^{y}$ will be 1 if any IDC module is constructed until year *y* as shown by (2.23) and *Q* is a large scalar. The requests received by each UG in each scenario are equal to the requests that are transferred to the IDC and the requests that are transferred to other UGs as shown in (2.24). The requests exchanged between the UGs are directional, as shown in (2.25). The proposed master problem (2.16)-(2.25), is an MIP problem in which the binary variables represent the decisions on the installation of IDC modules and data routes, the integer variables determine the number of active servers to serve the customers' request, and the continuous variables represent the rate of requests processed by the IDC modules. The procured solution is passed to the feasibility check sub-problem as described in the next section.

### 2.2.3 Electricity Network Security Check Sub-problem

The network security check sub-problem is formulated as (2.26)-(2.31). The objective is to minimize the mismatch between the generation and demand for all buses at each period and scenario, considering the imposed electricity demand by the IDCs. The nodal power balance for each period and scenario in the electricity network is shown in (2.27). The generation dispatch of thermal and wind units in each period and scenario is limited to their capacity as shown in (2.28) and (2.29). The unavailability of generation unit is captured by a corresponding binary parameter representing the state of the unit. The dc power flow formulation in which the flow of the line is dependent on the difference between the voltage angles at the ends of the line is presented in (30). If a transmission line is unavailable ($U_l^{p,y,s} = 0$) the voltage angles at the ends of the line are relaxed. The flow of the line is further limited by (31). If there is a mismatch in (2.26), the positive slack variables will be non-zero and the infeasibility Benders cut (2.32) is generated and sent to the master problem (2.16)-(2.25). Here $\widehat{W}^{p,y,s}$ is the value of the objective function. This iterative process continues until there is no violation in the sub-problem and the slack variables are zero.

$$\min W^{p,y,s} = \sum_b \left(t_{1,b}^{p,y,s} + t_{2,b}^{p,y,s}\right) \tag{2.26}$$

$$s.t.\ \sum_{\phi\in\psi_\phi^b} P_\phi^{p,y,s} + \sum_{w\in\psi_w^b} P_w^{p,y,s} - \sum_{d\in\psi_d^b} P_d^{p,y,s} - \sum_{i\psi_i^b}\begin{bmatrix}\sum_n \hat{m}_{i,n}^{p,y,s}\cdot[P_{idle}+P_{peak}]\\ +(P_{peak}-P_{idle})\times\mu^{-1}\cdot\sum_j\sum_n \hat{\lambda}_{j,i,n}^{p,y,s}\end{bmatrix}$$

$$+t_{1,b}^{p,y,s} - t_{2,b}^{p,y,s} = \left[\sum_{l\in L_{t,b}} PL_l^{p,y,s} - \sum_{l\in L_{f,b}} PL_l^{p,y,s}\right] \qquad (\tau_b^{p,y,s}) \tag{2.27}$$

$$0 \le P_\phi^{p,y,s} \le P_\phi^{max}\cdot U_\phi^{p,y,s} \tag{2.28}$$

$$0 \le P_w^{p,y,s} \le P_w^{f,p,y}\cdot U_w^{p,y,s} \tag{2.29}$$

$$-Q\cdot\left(1-U_l^{p,y,s}\right) \le X_{b,a}^{-1}\cdot\left[PL_l^{p,y,s} - \left(\theta_b^{p,y,s} - \theta_a^{p,y,s}\right)\right] \le Q\cdot(1-U_l^{p,y,s}) \tag{2.30}$$

$$\left|PL_l^{p,y,s}\right| \le PL_l^{max}\cdot U_l^{p,y,s} \tag{2.31}$$

$$\widehat{W}^{p,y,s} + \sum_b\sum_{i\in\psi_i^b}\tau_b^{p,y,s}\times\begin{bmatrix}(P_{peak}-P_{idle})\cdot\mu^{-1}\cdot\sum_j\sum_n\left(\lambda_{i,j,n}^{p,y,s}-\hat{\lambda}_{i,j,n}^{p,y,s}\right)\times\\ (P_{idle}+P_{peak})\cdot\left(\sum_n m_{i,n}^{p,y,s}-\hat{m}_{i,n}^{p,y,s}\right)\end{bmatrix} \le 0 \tag{2.32}$$

### 2.2.4 ISO's Optimal Operation Problem

Once the feasibility of the electricity network is ensured, the solution is passed to the ISO to determine the LMP of energy in the electricity network. This problem addresses the economic dispatch in the electricity network considering the imposed demand by the IDCs. The objective function is the operation cost of the generation units for each scenario as shown in (2.33) that is subjected to nodal generation and load balance (2.34), and other constraints of the electricity network (2.28)-(2.31). The Lagrange multiplier associated with nodal generation demand balance determines the LMP of electricity in each scenario as shown in (2.34).

$$\min \sum_s\sum_y\sum_p\sum_\phi F_c(P_\phi^{p,y,s}) \tag{2.33}$$

$$s.t.\ \sum_{\phi\in\psi_\phi^b} P_\phi^{p,y,s} + \sum_{w\in\psi_w^b} P_w^{p,y,s} - \sum_{d\in\psi_d^b} P_d^{p,y,s} - \sum_{i\psi_i^b}\begin{bmatrix}\sum_n \hat{m}_{i,n}^{p,y,s}\cdot[P_{idle}+P_{peak}]\\ +(P_{peak}-P_{idle})\cdot\mu^{-1}\cdot\sum_j\sum_n \hat{\lambda}_{j,i,n}^{p,y,s}\end{bmatrix}$$

$$= \sum_{l\in L_{t,b}} PL_l^{p,y,s} - \sum_{l\in L_{f,b}} PL_l^{p,y,s} \qquad (\delta_b^{p,y,s}) \tag{2.34}$$

$$|1 - Z^{new}/Z^{old}| \le \epsilon \tag{2.35}$$

Once the LMPs of electricity are updated, the master problem in (2.17) is solved and the operation cost of the IDC and the decisions made in the master problem is updated accordingly. The updated solution

of the master problem is passed to the security check sub-problem to check for the feasibility of the decisions in the electricity network. Similarly, the LMPs of electricity will be updated with the updated decisions for installing the IDCs in the master problem. This iterative process could stop once the sum of the investment and operation costs of the IDCs and data routes is not changing as shown in (2.35) where ε is a small scalar. If the process does not converge in practice, the number of iterations could be limited by the number of interactions among the ISO(s) and DCI.

## 2.3 Case Study

In this section, a 6-bus power system and the IEEE 118-bus power system are used to evaluate the efficiency of the presented expansion-planning framework. The planning horizon is 20 years and the annual request growth rate in the data network is 10% [2.35]. The bandwidth of the fiber optics data routes is typically 1-6 Gbyte/sec. Here, only one type of data route with 4 Gbyte/sec bandwidths is considered [2.36], [2.37]. For the sake of simplicity, it is assumed that each request corresponds to 500 Kbps data transmission [2.38]. The cost of developing data routes to transmit 8,000 requests per second is 12,800 $/mile [2.39]. The installation cost of an IDC depends on several factors including the available space, capacity, cooling and the leveraged network technologies [2.40]. The average idle power consumption for each server is 100W ($P_{idle}$ = 100 W), the peak power consumption for a server is 200W ($P_{peak}$ = 200 W) [2.41], and the number of hours in each period is 730. The expansion planning of IDC in the data network, the electricity network security check sub-problem, and the economic dispatch are solved using Gurobi 5.6. Scenario generation and reduction techniques to address the uncertainties are presented in the next section.

### 2.3.1 Scenario Generation and Reduction

The stochastic solution captures the uncertainties in the rate of requests in the data network, electricity demand, wind generation, and the outages in generation and transmission components in the power network. A large number of scenarios was generated using Monte-Carlo simulation and scenario reduction techniques were utilized to reduce the number of effective scenarios by eliminating the low probability scenarios and by bundling the comparable scenarios [2.43]. The forecast error in electricity demand and rate of requests for UGs are presented by Gaussian distribution function with the mean equal to the forecasted

electrical demand and rate of request for UG in each period and the standard deviations equal to 3% of the mean values. The Weibull probability distribution function is used to represent the uncertainty in the wind speed [2.44], [2.45] and wind generation is determined using the wind speed and wind turbine speed-power curve [2.46]. In this case, the uncertainties are captured by generating 3,000 scenarios and scenario reduction techniques including fast backward method, fast backward/forward method, and fast backward/backward method can be used to reduce the number of effective scenarios [2.43], [2.47]. These methods have diverse computational performance, and the choice of approach depends on the size of the problem and the expected level of solution accuracy. The fast-backward method provides the best computational performance with least accuracy for large scenario trees; however, the fast forward method procures more accurate results at the cost of longer computational time. In this chapter, the fast backward/forward method is selected to reduce the number of scenarios to 13.

### 2.3.2 6-Bus Power System

Figure 2-3 shows a sample 6-bus power system which represents the electricity network in this case. The characteristics of the generation units and transmission lines are shown in Tables 2-1 and 2-2, respectively.

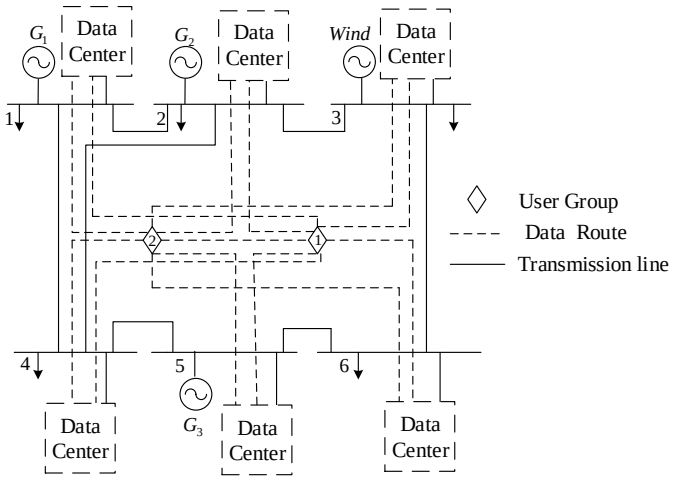


Figure 2-3 A 6-bus power system with candidate IDCs and data routes

Wind generation is located on bus 3. The annual demand growth in this network is 5%. The wind generation capacity is 18 MW. Here, the shape and scale parameter associated with Weibull distribution is set to 2.67 and 7.85, respectively [2.48].

Table 2-1 Thermal units' characteristics

| Unit | a ($/MW$^2$h) | b ($/MWh) | c ($/h) | $P_{max}$ (MW) |
|---|---|---|---|---|
| G1 | 0.79 | 6.6 | 1.4 | 40 |
| G2 | 1.1 | 7.6 | 1.3 | 30 |
| G3 | 2.5 | 10 | 1.1 | 20 |

Table 2-2 Transmission lines characteristics

| Line | From Bus | To Bus | Impedance (Ω) | Maximum Power Flow (MW) |
|---|---|---|---|---|
| 1 | 1 | 2 | 0.17 | 26 |
| 2 | 1 | 4 | 0.258 | 32 |
| 3 | 2 | 4 | 0.197 | 16 |
| 4 | 5 | 6 | 0.14 | 16 |
| 5 | 3 | 6 | 0.018 | 30 |
| 6 | 2 | 3 | 0.037 | 36 |
| 7 | 4 | 5 | 0.037 | 26 |

Table 2-3 Distances of UGs from buses in miles

| | B1 | B2 | B3 | B4 | B5 | B6 |
|---|---|---|---|---|---|---|
| $UG_1$ | 35 | 22 | 23 | 35 | 27 | 21 |
| $UG_2$ | 26 | 21 | 35 | 23 | 23 | 33 |

Table 2-4 Rate of requests for UGs in each period for the first year

| Period | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $UG_1$ | 3914 | 5548 | 4698 | 6875 | 6770 | 7695 |
| $UG_2$ | 4261 | 3701 | 4180 | 5063 | 7551 | 5822 |
| Period | 7 | 8 | 9 | 10 | 11 | 12 |
| $UG_1$ | 9182 | 6024 | 3790 | 7698 | 3757 | 1947 |
| $UG_2$ | 4156 | 2843 | 3775 | 7886 | 5806 | 5834 |

The data network consists of two UGs that receive the requests from the end users. All buses in the electricity network were considered as candidates for installing IDC modules. Each server in an IDC will process 3 requests per second. Three types of IDC modules, $k_1$, $k_2$, and $k_3$, with 500, 1000, and 1500 servers were considered respectively. The total power consumption of a data center module type $k_1$, $k_2$, and $k_3$ considering full CPU utilization is 0.2, 0.4, and 0.6 MW and the installation cost is \$0.4M, \$0.64M, and \$0.96M respectively. The length of the data routes between the UGs and IDCs are presented in Table 2-3. The distance between the UGs is 15 miles. The desired response time of IDCs to the received requests should not exceed 300 msec. For the sake of simplicity, the latency in data routes is ignored. The demand rate for the UGs in each period in the first year is shown in Table 2-4. The expansion planning is performed for 20

years and each year consists of 12 equal periods, each consists of 730 hours. The following four cases are considered:

Case 1 – Deterministic solution without congestion

Case 2 – Deterministic solution with congestion

Case 3 – Deterministic solution with outages

Case 4 – Stochastic solution.

*1) Case 1-Deterministic Solution without Congestion:*

Tables 2-5 and 2-6 show the installation decisions for IDC modules and data routes, respectively.

Table 2-5 Expansion of IDCs in case 1

| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Bus 2 | - | $k_3$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ |
| Bus 3 | $k_3$ | $k_1$ | - | - | - | - | - | - | - | - |
| Bus 4 | $k_3$ | $k_1$ | - | - | - | - | - | - | - | - |
| Bus 6 | $k_1$ | $k_1$ | - | - | - | $k_2$ | $k_2$ | $k_2$ | - | - |
| Year | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Bus 2 | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_3$ | $k_3$ | $k_3$ | $k_3$ | $k_3$ | $k_3$ |
| Bus 3 | - | - | - | $k_1$ | - | - | - | - | - | - |

Table 2-6 Expansions of data routes in case 1

| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| ($UG_1$, bus 2) | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| ($UG_1$, bus 3) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ($UG_1$, bus 6) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ($UG_2$, bus 2) | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| ($UG_2$, bus 4) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Year | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| ($UG_1$, bus 2) | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| ($UG_2$, bus 2) | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |

The IDC modules were built on buses 2, 3, 4, and 6. The total number of installed modules is 29; including 5, 15, and 9 modules of type $k_1$, $k_2$, and $k_3$ respectively. One data route was installed between the UGs in year 1. As shown in Tables 2-5 and 2-6, buses that are closest to the UGs (i.e., buses 2, 3, and 6 that are close to $UG_1$, and buses 2 and 4 that are close to $UG_2$) were selected to install IDC modules. In year 1, the UGs are connected to exchange requests with each other. In period 4 of year 1, the rates of requests received by $UG_1$ and $UG_2$ are 6875, and 5063 requests per second respectively. In this period $UG_1$ passes 1753 requests per second to the $UG_2$, and $UG_2$ sends 6816 requests per second to the IDC at bus 4. In years 3-5 modules of type $k_2$ are installed at bus 2 and the rate of requests transferred from $UG_1$ to $UG_2$ increased until year 6. By

establishing modules of type $k_2$ at bus 6 in years 6-8, the exchange rate among the UGs decreases. The total cost for DCI, in this case, is $20.22M. The installation cost of IDC is $8.92M and the installation cost of data routes (IDR) is $1.93M, and the operational cost of data centers (OC) is $9.37M. Establishing the link between the UGs will save $470K in this case. In this case, the LMPs were updated 3 times by the ISO. The number of iterations for finding a feasible solution for the TOU in the first, second, and third iterations are 3, 1, and 1 respectively. The interaction between DCI and TOU is shown by measuring the mismatch in the objective function of the electricity network security check sub-problem in Figure 2-4.

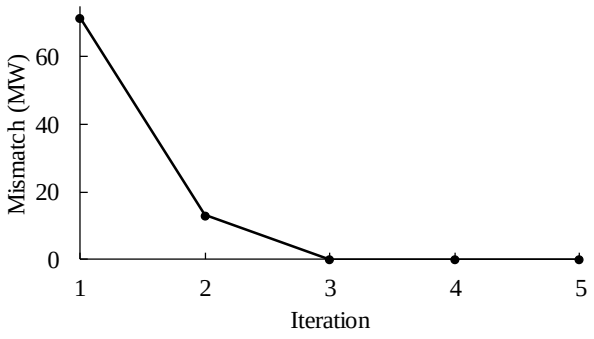


Figure 2-4 Mismatch in the electricity network security-check sub-problem at each iteration

Here, as the mismatch in the objective function of the sub-problem reduces to zero, the procured DCI's solution will become feasible in the electricity network. As shown in this figure, for the first iteration of updating the LMPs, TOU can serve the IDC modules after sending Benders cuts in three iterations. The solution time for this case is 7207 seconds.

*2) Case 2 – Deterministic Solution with Congestion:*

In this case, the capacity of lines "1" and "5" are reduced to 12 and 19 MW, respectively, to evaluate the effect of congestion in the power system. As a result, lines "1" and "5" are congested in years 15-20 and 17-20 respectively. The outcomes are shown in Tables 2-7 and 2-8. The total number of installed modules is 31; which includes 5, 22 and 4 modules of type $k_1$, $k_2$, and $k_3$, respectively. In this case buses 2, 3, 5, and 6 selected to install IDC modules. Comparing with Case 1, it is seen that the number of IDC modules installed at bus 3 and bus 2 was increased and decreased respectively. The total cost is increased by $890K, however,

installing the IDC modules at bus 3 relieves the congestion on line 5 and reduces the operation cost of the DCI by $30K. The total planning cost is increased to $21.11M. This cost includes the installation cost of the IDC ($9.24M), the IDR ($2.53M), and the OC ($9.34M). Comparing with Case 1, it is shown that the installation cost of IDC and IDR were increased, whereas the OC is decreased. In addition, $322K was saved in total expansion and operation cost because of establishing data routes between the UGs. The LMPs were updated 3 times by the ISO in this case. The number of iterations for finding a feasible solution in the first, second, and third iterations of updating the LMPs are 3, 5, and 4 respectively. The total solution time is 8025 seconds.

Table 2-7 Expansion of IDCs in case 2

| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Bus 2 | $k_3$ | $k_1$ | - | - | - | - | - | - | - | - |
| Bus 3 | $k_3$ | $k_1$ | - | - | $k_2$ | - | $k_2$ | $k_2$ | $k_2$ | $k_2$ |
| Bus 5 | $k_1$ | $k_1$ | - | - | - | - | - | - | - | - |
| Bus 6 | $k_1$ | $k_2$ | $k_2$ | $k_2$ | - | $k_2$ | - | - | - | - |
| Year | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Bus 2 | - | - | - | - | - | - | - | $k_2$ | - | $k_2$ |
| Bus 3 | $k_2$ | $k_2$ | $k_2$ | $k_3$ | $k_3$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ |
| Bus 5 | - | - | - | - | - | $k_2$ | - | - | - | - |
| Bus 6 | - | - | - | - | - | - | $k_2$ | - | $k_2$ | - |

Table 2-8 Expansion of data routes in case 2

| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| ($UG_1$, bus 3) | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 0 |
| ($UG_1$, bus 6) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ($UG_2$, bus 2) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ($UG_2$, bus 3) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ($UG_2$, bus 5) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ($UG_2$, bus 6) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Year | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| ($UG_1$, bus 3) | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| ($UG_2$, bus 2) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| ($UG_2$, bus 3) | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |

*3) Case 3 – Deterministic Solution with Contingencies:*

In this case, the contingencies because of failures in generation and transmission components were considered in the proposed framework. The forced outage rate (FOR) of transmission lines and generation units are 1% and 4%, respectively [2.49], [2.50]. The results are presented in Tables 2-9 and 2-10. By comparing this case with Case 1, it is shown that the IDC modules were installed on buses 1, and 5 instead of buses 2, and 4. In year 1, line "2" and generator G3 are on outage in periods 3 and 7 respectively. Therefore, modules of type $k_1$, and $k_3$ were installed on buses 1 and 5 instead of bus 4 to ensure the electricity supply security for the

IDC modules. As a result of failures of generator G2, and transmission line "1" in periods 2, and 11 in year 3, it is impossible to provide the required energy for the IDC modules located on bus 2. Therefore, compared to Case 1, IDC module of type k2 was installed at bus 5 instead of bus 2. The total cost of planning is \$20.81M; including the installation cost of the IDC (\$9.19M), the IDR (\$2.18M), and the OC (\$9.44M). Comparing with Case 1, the total cost is increased by \$590K and the OC of IDCs is increased by 5.82% because of outages in the electricity network. Ignoring the option of establishing data routes between $UG_1$ and $UG_2$ will increase the total cost to \$20.93M, which shows a \$120K increase in the total planning cost. The LMPs were updated 2 times by the ISO, and the number of iterations for finding a feasible solution in each iteration is one. The total solution time for this case is 5352 seconds. The total number of variables in the master problem for deterministic solutions, i.e., Case 1-3 is 7519 that includes 1440 integer variables, 800 binary variables, and 5279 continuous variables. The number of continuous variables in the electricity network security check sub-problem is 9601.

Table 2-9 Expansion of data centers in case 3

| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Bus 1 | $k_1$ | $k_1$ | - | - | - | - | - | - | - | - |
| Bus 3 | $k_1$ | $k_1$ | - | - | - | - | - | - | - | - |
| Bus 5 | $k_3$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ |
| Bus 6 | $k_3$ | $k_1$ | - | - | - | - | - | - | - | - |
| Year | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Bus 1 | - | - | - | $k_2$ | - | - | - | - | - | - |
| Bus 3 | - | - | - | - | $k_2$ | - | - | - | - | - |
| Bus 5 | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_3$ | $k_3$ | $k_3$ | $k_2$ | $k_2$ |
| Bus 6 | - | - | - | - | - | - | - | - | $k_2$ | $k_2$ |

Table 2-10 Expansion of data routes in case 3

| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| ($UG_1$, bus 3) | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ($UG_1$, bus 5) | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| ($UG_1$, bus 6) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ($UG_2$, bus 1) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ($UG_2$, bus 5) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Year | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| ($UG_1$, bus 5) | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 |
| ($UG_2$, bus 5) | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| ($UG_1$, bus 6) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 |

*4) Case4 – Stochastic Solution:*

In this case, the uncertainties in the demand, renewable generation, and availability of generation and transmission resources as well as the volatility in the rate of requests received by the UGs were considered.

Using scenario reduction techniques, 13 scenarios with the probabilities shown in Table 2-11 were considered. Tables 2-12 and 2-13 show the outcomes of the expansion planning process. In this case, 37 IDC modules are installed including; 12, 21, and 4 modules of type $k_1$, $k_2$, and $k_3$ respectively.

Table 2-11 Probability of scenarios

| Scenario | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Probability (%) | 88.8 | 0.2 | 0.2 | 2 | 2.2 | 2 | 0.2 |
| Scenario | 8 | 9 | 10 | 11 | 12 | 13 | |
| Probability (%) | 0.2 | 0.2 | 1.8 | 0.2 | 1.8 | 0.2 | |

Table 2-12 Expansion of IDCs in case 4

| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Bus1 | $k_1$ | - | - | - | - | - | - | - | - | - |
| Bus2 | $k_3$ | $k_1$ | $k_1$ | - | $k_2$ | $k_2$ | - | - | - | - |
| Bus3 | $k_3$ | $k_1$ | - | - | - | - | - | - | - | - |
| Bus4 | $k_1$ | $k_2$ | $k_2$ | $k_2$ | - | - | $k_3$ | $k_2$ | $k_2$ | $k_2$ |
| Bus6 | $k_1$ | $k_1$ | - | - | - | - | - | - | - | - |
| Year | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Bus1 | - | - | - | $k_1$ | $k_1$ | - | $k_2$ | - | - | - |
| Bus3 | - | - | - | - | - | - | - | - | $k_2$ | $k_2$ |
| Bus4 | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_2$ | $k_3$ | $k_2$ | $k_2$ |
| Bus5 | - | - | - | - | - | $k_2$ | $k_1$ | - | - | - |
| Bus6 | - | - | - | $k_1$ | $k_1$ | - | - | - | - | - |

Table 2-13 Expansion of data routes in case 4

| Year | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| ($UG_1$, bus 1) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ($UG_1$, bus 3) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ($UG_1$, bus 4) | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| ($UG_1$, bus 6) | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| ($UG_2$, bus 2) | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| ($UG_2$, bus 4) | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| Year | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| ($UG_1$, bus 1) | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| ($UG_1$, bus 3) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 |
| ($UG_1$, bus 4) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| ($UG_2$, bus 4) | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 |
| ($UG_2$, bus 5) | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |

The IDC modules are installed on all buses. The UGs are connected in years 1, and 16. The number of data routes installed between $UG_1$ and the IDC modules located on buses 1, 3, 4, and 6 is 2, 3, 3, and 1 respectively. Similarly, 2, 5, and 1 data routes are installed between $UG_2$ and IDC modules located at buses 2, 4, and 5 respectively. In years 1 and 2, $UG_1$ redirects part of the received requests to $UG_2$ to exploit the capacity of installed data routes between $UG_2$ and the IDC modules located on buses 2, and 4. In years 3-6, the IDC modules were installed on buses 2 and 4, where part of the requests of $UG_1$ is redirected to the $UG_2$. In year 16, the IDC modules were installed on buses 4 and 5 and another data route is established between the UGs to accommodate the increase in the rate of requests sent from $UG_1$ to $UG_2$. The total cost of planning for this

case is increased to $22.84M compared to Case 1. The installation cost of the IDC is $10.31M; the IDR is $3.15M and the expected OC is $9.38M. The total number of variables in the master problem for this case is 88159 that includes 18720 integer variables, 800 binary variables, and 68639 continuous variables. The number of continuous variables in the electricity network security check sub-problem is 124801. The LMPs are updated by the ISO in two iterations. In the first and second iterations, it takes 5 and 1 iterations to reach a feasible solution for the TOU respectively. The solution time is 8950 seconds.

### 2.3.3 IEEE 118-Bus System

This case captures the IEEE 118-bus system with four UGs receiving the data requests from the end-users in the data network. In this case study, it is assumed that only one type of IDC module could be installed annually. A module in an IDC consists of 3500 servers (M=3500). Hence, the total power consumption of an IDC module considering full CPU utilization is 1.4MW. The installation cost of the IDC module with such demand is $2.8M [2.42]. The candidates for installing IDC modules are 23 buses in the electricity network. Each server in an IDC module will process 2 requests per second and the desired response time of the IDC modules is less than 300 msec. The annual growth rate of electrical demand is 3%. There are five wind generation units, each with the generation capacity of 150 MW. Here, the shape and scale parameters associated with Weibull distribution are set to 2.1 and 8.3, respectively. The rates of requests received by UGs in each period of the first year are shown in Table 2-14. The expansion planning is performed for 20 years and each year consists of 4 equal periods each consists of 2190 hours.

Table 2-14 Rate of requests received by UGs (request per second)

| Period | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $UG_1$ | 6,954 | 14,476 | 20,840 | 24,630 |
| $UG_2$ | 20,655 | 29,531 | 13,891 | 15,864 |
| $UG_3$ | 7,518 | 15,428 | 20,459 | 25,767 |
| $UG_4$ | 7,163 | 13,861 | 20,343 | 24,721 |

*1) Case 1- Deterministic Solution Without Congestion:*

In this case, 119 IDC modules were installed on buses 12, 17, 32, 37, 46, 49, 54, 56, 59, 62, 68, 69, 70, 80, 82, 86, 89, 91, 100 and 105. Moreover, 87 data routes were installed between UGs and IDCs 3 data routes were installed between $UG_1$ and $UG_2$, 3 data routes were installed between $UG_2$ and $UG_3$, and 1 data route

was installed between $UG_1$ and $UG_4$. The LMPs were updated 3 times by the ISO in this case. The number of iterations for finding a feasible solution in the first, second, and third iterations is 5, 3, and 4 respectively. Figure 2-5 shows the number of interactions between the DCI and TOU to ensure the energy security for the installed IDC modules in first iteration of updating the LMPs. As shown in this figure, the mismatch in the electricity network security-check sub-problem reached zero after 5 iterations. This means that after 5 iterations, the decisions made by the DCI is feasible in the electricity network. The solution time for this case is 12558 seconds.

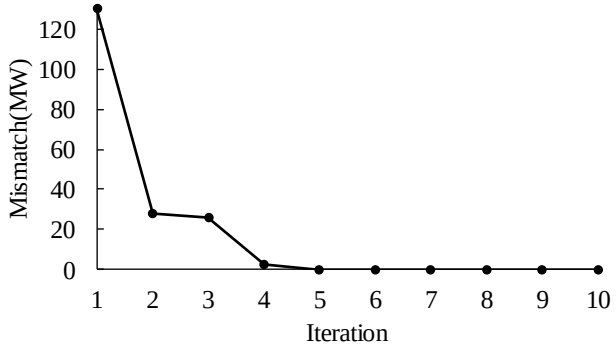


Figure 2-5 Mismatch in the electricity network security-check sub-problem at each iteration

*2) Case2 - Deterministic Solution with Congestion:*

In this case the capacity of the transmission lines 47, 48, 50-53, 62-64, 74-80, 82, 84-86, 104, 107, 158-160, 163,164 and 183 are reduced. In this case, 117 IDC modules were installed on buses 5, 17, 32, 37, 46, 49, 54, 56, 59, 62, 68, 69, 70, 80, 82, 86, 89, 91, 100 and 105. As a result of congestion, the number of modules installed on the buses is changed compared to Case 1. For example, the installed IDC modules on buses 37 and 56 are reduced from 16 and 15, to 4 and 8, respectively. As a result of congestion, the total cost is increased by \$3.41M. The IDR and the installation cost of the IDC modules are increased by \$6.52M and \$570K respectively. The OC is decreased by \$3.68M. The number of installed data routes between $UG_1$ and $UG_2$, $UG_2$ and $UG_3$, and $UG_2$ and $UG_4$ are 2, 2 and 2, respectively. The LMPs were updated 3 times by the ISO in this case. The number of iterations for finding a feasible solution in the first, second, and third iterations of updating the LMPs is 5, 2, and 3 respectively. The solution time for this case is 12407 seconds.

*3) Case3 – Deterministic Solution with Outages:*

The FOR of transmission lines and generation units for this case is 1% and 4%, respectively. In this case, 117 IDC modules were installed at buses 12, 17, 23, 25, 32,37, 46, 49, 54, 56, 59, 62, 68, 69, 70, 80, 82, 86, 89, 100 and 105. The total cost is increased by $5.97M compared to Case 1. In this case, the OC, and IDR, were increased by 2.14%, and 15.23% respectively and the installation cost of the IDC was decreased by 1.1% compared to Case 1. In this case, 2, 3 and 3 data routes were installed between $UG_1$ and $UG_2$, $UG_2$ and $UG_3$ and $UG_2$ and $UG_4$, respectively. The LMPs were updated 3 times by the ISO in this case. The number of iterations for finding a feasible solution in the first, second, and third iteration of updating the LMPs is 5, 2, and 2 respectively. The solution time is 13455 seconds. The number of variables in the master problem of Cases 1-3 is 72931 that includes 9440 integer variables, 14480 binary variables, and 62043 continuous variables.

*4) Case4 – Stochastic Solution:*

In this case, using scenario reduction techniques, 13 scenarios with the probabilities shown in Table 2-15 were considered. Here, 134 IDC modules were constructed on buses 5, 12, 17, 23, 25, 37, 46, 49, 54, 56, 59, 62, 68, 69, 70, 80, 82, 86, 89 and 100. The total cost is increased by $31.8M, compared to Case 1.

Table 2-15 Probability of scenarios

| Scenario | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Probability (%) | 79.8 | 2.4 | 1.4 | 1.8 | 1.8 | 1.4 | 1.8 |
| Scenario | 8 | 9 | 10 | 11 | 12 | 13 | |
| Probability (%) | 1.8 | 1.6 | 1.8 | 1.4 | 1.4 | 1.6 | |

Table 2-16 Planning cost for All case 1-4

| | Installation cost of IDC (M$) | IDR (M$) | OC (M$) | Total Cost (M$) |
|---|---|---|---|---|
| Case 1 | 135.84 | 27.44 | 153.72 | 317 |
| Case 2 | 134.43 | 33.30 | 151.55 | 319.28 |
| Case 3 | 134.34 | 31.62 | 157.01 | 322.97 |
| Case 4 | 170.16 | 36.40 | 142.24 | 348.80 |

The total number of data routes is 91 and the number of data routes installed between the UGs is 8. There are 3, 3 and 2 data routes installed between $UG_1$ and $UG_2$, $UG_2$ and $UG_3$, and $UG_2$ and $UG_4$, respectively. The outcomes of the expansion planning for all cases are summarized in Table 2-16. The LMPs were updated by the ISO in 2 iterations. The number of iterations for finding a feasible solution in the first, and second iteration

of updating the LMPs is 5, and 2 respectively. The solution time for this case is 26307 seconds. The number of variables in the master problem is 767971 that includes 122720 integer variables, 14480 binary variables, and 630751 continuous variables.

## 2.4 Summary

In this chapter, a coordinated expansion-planning framework for IDCs in electricity and data networks is presented. The proposed expansion planning formulation and solution methodology capture the uncertainties in the generation and transmission in the electricity network as well as the uncertainty of demand in electricity and data networks. Scenario based stochastic programming is used to solve the proposed problem and Benders decomposition is used to capture the interaction between the decision makers in the electricity network, i.e., TOU, and the investor in the data network, i.e., DCI. The objective is to minimize the installation and operation costs of IDCs and data routes in the electricity and data networks. The effect of congestion and unavailability of generation and transmission components in the electricity network on the expansion planning decisions in the data network was addressed in the case studies. It is shown that the interaction among the TOU, ISO and DCI improves the economics and energy supply security for the IDC modules. The congestion in the electricity network would change the expansion decisions in the data network because of the energy supply deficiency.

## 2.5 References


[2.1] J. Rath, “Data center knowledge, data center strategies, simplifying high stakes, mission critical decisions in a complex industry,” Vantage Data Centers, Rath Consulting, Saskatoon, SK, Canada, Tech. Rep., Jul. 2011.

[2.2] J. Koomey, *Growth in Data Center Electricity Use*. Burlingame, CA, USA: Analytics Press, 2011.

[2.3] J. Hamilton, “Cooperative expendable micro-slice servers (CEMS): Low cost, low power servers for Internet-scale services,” in *Proc. 4th Biennial Conf. Innov. Data Syst. Res. (CIDR),* Jan. 2009, pp. 770–791.

[2.4] D. Brooks and M. Martonosi, “Dynamic thermal management for highperformance microprocessors,” in *Proc. 7th Int. Symp. High Perform. Comput. Archit.*, Nuevo León, Mexico, 2001, pp. 171–182.

[2.5] S. Nedevschi, L. Popa, G. Iannaccone, S. Ratnasamy, and D. Wetherall, “Reducing network energy consumption via sleeping and rateadaptation,” in *Proc. 5th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, San Francisco, CA, USA, Apr. 2008, pp. 323–336.

[2.6] W. Deng, F. Liu, H. Jin, C. Wu, and X. Liu, “MultiGreen: Costminimizing multi-source datacenter power supply with online control,” in *Proc. 4th Int. Conf. Future Energy Syst., Berkeley*, CA, USA, 2013, pp. 149–160.

[2.7] E. V. Carrera, E. Pinheiro, and R. Bianchini, “Conserving disk energy in network servers,” in *Proc. 17th Int. Conf. Supercomput.*, San Francisco, CA, USA, 2003, pp. 86–97.

[2.8] C. Clark et al., “Live migration of virtual machines,” *in Proc. USENIX Conf. Symp. Netw. Syst. Design Implement. (NSDI)*, Boston, MA, USA, 2005, pp. 273–286.

[2.9] *NIST Framework and Roadmap for Smart Grid Interoperability Standards*, Nat. Inst. Stand. Technol., Gaithersburg, MD, USA, 2010.

[2.10] *National Assessment of Demand Response Potential*, Federal Energy Regul. Commission, Washington, DC, USA, 2009.

[2.11] Z. Liu, I. Liu, S. Low, and A. Wierman, “Pricing data center demand response,” in *Proc. ACM SIGMETRICS*, Austin, TX, USA, 2014, pp. 111–123.

[2.12] Z. Zhou, F. Liu, Z. Li, and H. Jin, “When smart grid meets geodistributed cloud: An auction approach to datacenter demand response,” in *Proc. IEEE INFOCOM*, Hong Kong, Apr./May 2015, pp. 2650–2658.

[2.13] Z. Zhou, F. Liu, and Z. Li, “Pricing bilateral electricity trade between smart grids and hybrid green datacenters,” in *Proc. ACM Int. Conf. Meas. Model. Comput. Syst. (SIGMETRICS)*, 2015, pp. 1–2.

[2.14] Z. Liu, A. Wierman, Y. Chen, B. Razon, and N. Chen, “Data center demand response: Avoiding the coincident peak via workload shifting and local generation,” in *Proc.*

*SIGMETRICS Int. Conf. Meas. Model. Comput. Syst.*, Pittsburgh, PA, USA, 2013, pp. 341–342.

[2.15] Z. Zhou et al., "Fuel cell generation in geo-distributed cloud services: A quantitative study," in *Proc. IEEE ICDCS,* Madrid, Spain, 2014, pp. 52–61.

[2.16] J. Camacho, Y. Zhang, M. Chen, and D. M. Chiu, "Balance your bids before your bits: The economics of geographic load-balancing," in *Proc. 5th Int. Conf. Future Energy Syst. (e-Energy),* Cambridge, U.K., Jun. 2014, pp. 75–85.

[2.17] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. H. Andrew, "Geographical load balancing with renewables," *ACM SIGMETRICS Perform. Eval. Rev.,* vol. 39, no. 3, pp. 62–66, 2011.

[2.18] J. Li, Z. Bao, and Z. Li, "Modeling demand response capability by Internet data centers processing batch computing jobs," *IEEE Trans. Smart Grid*, vol. 6, no. 2, pp. 737–747, Mar. 2015.

[2.19] J. Heo et al., "OptiTuner: On performance composition and server farm energy minimization application," *IEEE Trans. Parallel Distrib. Syst*., vol. 22, no. 11, pp. 1871–1878, Nov. 2011.

[2.20] Y. Yao, L. Huang, A. Sharma, L. Golubchik, and M. Neely, "Data centers power reduction: A two time scale approach for delay tolerant workloads," in *Proc. INFOCOM*, Orlando, FL, USA, 2012, pp. 1431–1439.

[2.21] K. Wang, M. Lin, F. Ciucu, A. Wierman, and C. Lin, "Characterizing the impact of the workload on the value of dynamic resizing in data centers," *SIGMETRICS Perform. Eval. Rev.*, vol. 40, no. 1, pp. 405–406, Jun. 2012.

[2.22] M. Ghamkhari and H. Mohsenian-Rad, "Optimal integration of renewable energy resources in data centers with behind-the-meter renewable generator," in *Proc. IEEE Int. Conf. Commun*., Ottawa, ON, Canada, Jun. 2012, pp. 3340–3344.

[2.23] S. Salinas, M. Li, P. Li, and Y. Fu, "Dynamic energy management for the smart grid with distributed energy resources," *IEEE Trans. Smart Grid,* vol. 4, no. 4, pp. 2139–2151, Dec. 2013.

[2.24] A. Wierman, Z. Liu, I. Liu, and H. M. Rad, “Opportunities and challenges for data center demand response,” in *Proc. IEEE Int. Conf. Green Comput. Conf. (IGCC)*, 2014, pp. 1–10.

[2.25] Z. Zhou et al., “Carbon-aware load balancing for geo-distributed cloud services,” in *Proc. IEEE Comput. Soc. Annu. Int. Symp. Model. Anal. Simulat. Comput. Telecommun. Syst. (MASCOTS),* San Francisco, CA, USA, 2013, pp. 232–241.

[2.26] A. Vafamehr and M. E. Khodayar, “Expansion planning of data centers in energy and cyber networks,” *in Proc. IEEE PES Gen. Meeting*, Boston, MA, USA, 2016, pp. 1–5.

[2.27] A. Rahman, X. Liu, and F. Kong, “A survey on geographic load balancing based data center power management in the smart grid environment,” *IEEE Commun. Surveys Tuts,* vol. 16, no. 1, pp. 214–233, 1st Quart, 2013.

[2.28] *Business Growth, Storage Drive Datacenter Expansion.* Accessed: Aug. 2017. [Online]. Available: http://www.enterprisetech.com/2014/10/ 24/business-growth-storage-drive-datacenter-expansion/

[2.29] C. Li *et al*., “Enabling datacenter servers to scale out economically and sustainably,” in *Proc. 46th Annu. IEEE/ACM Int. Symp. Microarchitect. (MICRO),* 2013, pp. 322–333.

[2.30] M. Shahidehpour and Y. Fu, “Benders decomposition: Applying benders decomposition to power systems,” *IEEE Power Energy Mag.*, vol. 3, no. 2, pp. 20–21, Mar./Apr. 2005.

[2.31] X. Fan, W. Weber, and L. A. Barroso, “Power provisioning for a warehouse-sized computer,” in *Proc. 34th Annu. Int. Symp. Comput. Architect,* San Diego, CA, USA, 2007, pp. 13–23.

[2.32] *NGiNx.* Accessed: Aug. 2017. [Online]. Available: https:// www.nginx.com/resources/deployment-guides/load-balance-node-js/

[2.33] C. Kelley and J. Cooley, *Deploying and Using Containerized/Modular Data Center Facilities,* Green Grid, Beaverton, OR, USA, 2011.

[2.34] L. Rao, X. Liu, L. Xie, and W. Liu, “Minimizing electricity cost: Optimization of distributed Internet data centers in a multi-electricitymarket environment,” in *Proc. IEEE INFOCOM*, San Diego, CA, USA, Mar. 2010, pp. 1–9.

[2.35] R. Smith, “Computing in the cloud,” *Res. Technol. Manag*., vol. 52, no. 5, pp. 65–68, 2009.

[2.36] *14 Tbps Over a Single Optical Fiber: Successful Demonstration of World’s Largest Capacity—145 Digital High-Definition Movies Transmitted in One Second.* Tokyo, Japan: NTT Press Release, Sep. 2006.

[2.37] M. S. Alfiad et al., “111-Gb/s POLMUX-RZ-DQPSK transmission over 1140 km of SSMF with 10.7-Gb/s NRZ-OOK neighbours,” in *Proc. ECOC,* Brussels, Belgium, 2008, pp. 1–2.

[2.38] *Netflix Viewing and Data Usage.* Accessed: Aug. 2017. [Online]. Available: http://www.nbnco.com.au/blog/entertainment/how-muchdata-does-streaming-video-movies-and-tv-use.html

[2.39] *U.S. Department of Transportation.* Accessed: Aug. 2017. [Online]. Available: http://www.itscosts.its.dot.gov/its/benecost.nsf/DisplayRUC ByUnitCostElementUnadjusted?ReadForm&UnitCostElement=Fiber+ Optic+Cable+Installation+&Subsystem=Roadside+Telecommunica tions+

[2.40] *Online Tech*. Accessed Aug. 2017. [Online]. Available: http://www.onlinetech.com/resources/references/leasing-vs-building-adata-center

[2.41] M. Ghamkhari and H. Mohsenian-Rad, “Profit maximization and power management of green data centers supporting multiple slas,” in *Proc. IEEE Int. Conf. Comput. Netw. Commun. (ICNC),* San Diego, CA, USA, 2013, pp. 465–469.

[2.42] *Data Center Solutions.* Accessed: Aug. 2017. [Online]. Available: http://www.mdiaccess.com/portfolio/

[2.43] J. Dupacová, N. Gröwe-Kuska, and W. Römisch, “Scenario reduction in ˇ stochastic programming: An approach using probability metrics,” *Math. Program.* vol. 95, pp. 493–511, Jan. 2003.

[2.44] *Wind Power Program.* Accessed: Aug. 2017. [Online]. Available: http:// www.wind-power-program.com/wind_statistics.htm

[2.45] J. Waewsak, C. Chancham, M. Landry, and Y. Gagnon, “An analysis of wind speed distribution at Thasala, Nakhon Si Thammarat, Thailand,” *J. Sustain. Energy Environ,* vol. 2, no. 2, pp. 51–55, Apr./Jun. 2011.

[2.46] G. A. Orfanos, P. S. Georgilakis, and N. D. Hatziargyriou, “Transmission expansion planning of systems with increasing wind power integration*,” IEEE Trans. Power Syst*., vol. 28, no. 2, pp. 1355–1362, May 2013.

[2.47] *GAMS/SCENRED Documentation*. Accessed: Aug. 2017. [Online]. Available: https://www.gams.com/latest/docs/T_SCENRED.html

[2.48] D. Indhumathy, C. V. Seshaiah, and K. Sukkiramathi, “Estimation of Weibull parameters for wind speed calculation at Kanyakumari in India,” *Int. J. Innov. Res. Sci. Eng. Technol.,* vol. 3, pp. 8340–8345, Jan. 2014

[2.49] D. O. Koval and A. A. Chowdhury, “Assessment of transmission-line common-mode, station-originated, and fault-type forced-outage rates,” *IEEE Trans. Ind. Appl.*, vol. 46, no. 1, pp. 313–318, Jan./Feb. 2010.

[2.50] Muskrat Falls Project, *Forced Outage Rates 2006 Update*, Syst. Plan. Hydro Power Commitment, Ottawa, ON, Canada, Dec. 2006, accessed: Aug. 2017. [Online]. Available: http://www.science.smith.edu/~jcardell/ Courses/EGR325/Readings/FOR%202006Update.pdf

[2.51] A. Vafamehr and M. E. Khodayar. Energy-aware cloud computing. The Electricity Journal, Elsevier, 31(2): 40-49 March 2018.

# Chapter 3

## NETWORK CONSTRAINED EXPANSION PLANNING OF BATTERY ENERGY STORAGES IN DISTRIBUTION NETWORK WITH DATA CENTERS

This chapter proposes a framework for expansion planning of BES in the distribution network with PV solar generation that captures the interaction among the DCOs and the DSO. In the presented framework the uncertainties associated with PV generation and demand assets; as well as the availability of distribution network components are considered. The proposed framework minimizes the installation and operational cost of the distribution network while ensuring the security and reliability of the distribution network as well as the quality of service for the end users in the data center. The contributions of the presented research are as follows:

- The expansion planning of BES in distribution network is addressed considering the flexible data center demands.
- The quality of service in data network is captured by modeling the DCOs.
- The interactions among the DSO and the DCOs are considered to ensure limited data sharing among the electricity and data network components.
- A dissimilarity-based sparse subset selection (DS3) algorithm is leveraged to cluster the hourly data set to limited data subsets that represent the hourly data for the expansion planning study.
- The uncertainties in the planning horizon are considered by capturing the variability in data network workload, electrical load and PV generation.

The structure of this chapter is as follows: in Section 3.1, the problem formulation is presented. In Sections 3.2 and 3.3, the solution methodology and a case study to show the effectiveness of the proposed framework are presented respectively. Finally, Section 3.4 presents the summary.

## 3.1 List of Symbols.

**Indices:**

| | |
|---|---|
| $a, b$ | Index of bus |
| $d$ | Index of electric load |
| $f$ | Index of main distribution feeder |
| $i$ | Index of distributed generation unit |
| $l$ | Index of electricity distribution line |
| $h$ | Index of hour |
| $t$ | Index of representative day |
| $y$ | Index of year |
| $s$ | Index of battery energy storage |
| $n$ | Index of data center |
| $pv$ | Index of photovoltaic unit |
| $sc$ | Index of scenario |

**Binary variables:**

| | |
|---|---|
| $k_s^{b,y}$ | Decision variable for installing battery energy storage $s$ of type $u$ |
| $I_s^{b,y}$ | Decision variable for installation of battery energy storage $s$ of type $u$ |

**Real variables:**

| | |
|---|---|
| $E_{s,sc}^{h,t,y}$ | Available energy in battery storage |
| $p_{f,b,sc}^{h,t,y} / q_{f,b,sc}^{h,t,y}$ | Exchanged real/reactive power from the main grid |
| $p_{i,sc}^{h,t,y} / q_{i,sc}^{h,t,y}$ | Real/reactive generation dispatch of distributed generation units |
| $p_{pv,sc}^{h,t,y} / q_{pv,sc}^{h,t,y}$ | Real/reactive power of photovoltaic unit |
| $p_{s,sc}^{h,t,y} / q_{s,sc}^{h,t,y}$ | Real/reactive dispatch of the battery energy storage |
| $p_{n,sc}^{h,t,y}$ | Power demand of data center |

| | |
|---|---|
| $p_{l,sc}^{h,t,y}/q_{l,sc}^{h,t,y}$ | Real/reactive power flow of line $l$ |
| $V_{b,sc}^{h,t,y}$ | Squared voltage magnitude at bus $b$ |
| $Z^{+}{}_{b,sc}^{h,t,y}, Z^{-}{}_{b,sc}^{h,t,y}$ | Slack variables capturing the real power mismatch (kW) |
| $Z'^{+}{}_{b,sc}^{h,t,y}, Z'^{-}{}_{b,sc}^{h,t,y}$ | Slack variables capturing the reactive power mismatch (kVar) |
| $Z^{\circ+}{}_{sc}^{h,t,y}, Z^{\circ-}{}_{sc}^{h,t,y}$ | Slack variables capturing the mismatch in the data network |
| $\Gamma_{n,sc}^{h,t,y}$ | Rate of requests allocated for each data center (requests/hour) |
| $\tau, \tau p, \tau q, \tau e, \pi$ | Lagrange Multipliers |

**Parameters:**

| | |
|---|---|
| $d_y$ | Annual discount rate |
| $IC_s$ | Investment costs of storage unit |
| $DT_t$ | Number of days in a representative |
| $NY$ | Total number of years in planning horizon |
| $\Psi_b$ | Set of potential locations for installing the battery energy storages |
| $\tilde{p}_{pv,sc}^{h,t,y}$ | Forecasted power of photovoltaic unit |
| $p_{d,sc}^{h,t,y}/q_{d,sc}^{h,t,y}$ | Real/reactive electricity demand at bus $b$ |
| $U_{l,sc}^{h,t,y}$ | Availability of the electric line $l$; 1 for being available and 0 otherwise |
| $r_l^{a,b}, x_l^{a,b}$ | Resistance/reactance of line $l$ between buses $a$ and $b$ |
| $S_l$ | Apparent power capacity of electricity lines |
| $\overline{V}_b, \underline{V}_b$ | Max/min squared voltage magnitude at bus $b$ |
| $\rho_{sc}$ | Probability of each scenario |
| $\Gamma_{tot}{}_{sc}^{h,t,y}$ | Total rate of requests received by the load balancer |
| $\alpha_i, \beta_i, \omega_i$ | Coefficients for the cost function of distributed generation units |
| $\gamma_{(.)}^{b}$ | Set of components connected to bus $b$ |
| $L_{f,b}/L_{t,b}$ | Set of lines starting from/ending to bus $b$ |
| $\eta_b$ | Set of buses connect the distribution network to the main utility grid |

| | |
|---|---|
| $\Gamma_n^{max}$ | Maximum requests that can be process in a data center |
| $p_n^{max}$ | Maximum power consumption of data center |
| $p_s^{max}/q_s^{max}$ | Real/reactive capacity of battery energy storage |
| $E_s^{max}$ | Capacity of the battery energy storage |
| $pr_{sc}^{h,t,y}$ | Price of electricity |
| $\phi_f$ | Allowable power factor angle at feeder |

## 3.2 Problem Description and Formulation

The proposed planning framework captures the interaction between the DCO and the DSO in the distribution network. The structure of the proposed framework is shown in Figure 3-1. In this figure, the DSO makes expansion decisions on the location, time, and type (size) of the BES units to minimize the incurred installation cost and operational cost of the distribution network. The DSO also checks for the feasibility of the procured decisions to ensure the security of the distribution network.

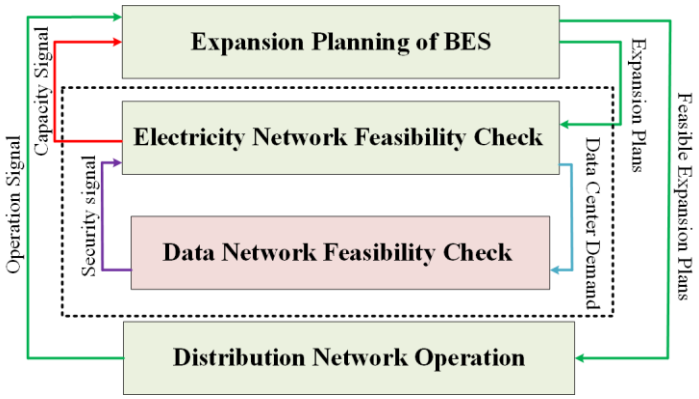


Figure 3-1 Expansion planning of battery energy storages

The DSO's decision also includes the dedicated supply for the data centers that is feasible in the distribution network. The dedicated supply is passed to DCO to check for feasibility of the provided energy to serve the required cloud computing services in the data network. In case the dedicated electric supply is not feasible for the data centers, DCO passes a security signal to the DSO to update the network feasibility check decisions. Once the dedicated supply for DCO is feasible for providing the cloud computing services, the capacity signal is passed from the electricity network feasibility check to the expansion planning stage.

Once the DSO's solution satisfies the electricity network constraints as well as the DCO's data network constraints, the expansion plans are used to update the operational cost of the distribution network. The operational signals would impose further constraints in the expansion planning of BES and may update the expansion plans accordingly. The interactions between the DCO and the DSO continues until the solution procured by the DSO is feasible in the data network and the operation cost is minimum in the electricity network.

In order to capture the interaction between the DCO and the DSO, Benders decomposition is used. Here the master problem captures the expansion planning of BES in the distribution network. The expansion decisions are passed to the electricity network feasibility sub-problem (sub-problem 1) in which the feasibility of the expansion decisions is verified. In this process, the considered demands for the data centers are passed to the DCO to evaluate for the feasibility of the supplied energy in the data network (sub-problem 2). Once data centers' power supply is feasible to serve the workloads in the data network, the infeasibility of the expansion decisions in the electricity network is reflected by Benders cuts to the master problem. If the expansion plans are feasible in the electricity network considering the data center demands for serving the workloads in the data network, the operational cost of the electricity network is calculated in the DSO optimality sub-problem and the operational cost signal is passed to the master problem to update the operational cost of the distribution network. The detailed problem formulation for the master problem and the sub-problems are presented below.

### 3.2.1 Expansion Planning Problem-Master Problem *(MP)*

The master problem is formulated in (3.1) -(3.4). Here, the expansion planning strategies are determined to minimize the installation cost of the BES systems. In (3.1), the lower bound of the solution is minimized. In (3.2), the lower bound of solution should be greater than or equal to the minimum installation cost of the BES systems considering the annual discount rate. The installation status of the BES systems satisfies (3.3), and (3.4) ensures that one type of storage unit will be installed at any potential location in the planning horizon.

$$\min Z_{lower} \tag{3.1}$$

$s.t.$

$$Z_{lower} \geq \sum_{y=1}^{NY}\left(1+d_y\right)^{1-y}.\left(\sum_{b\in\Psi_b}\sum_s IC_s.k_s^{b,y}\right) \quad (3.2)$$

$$I_s^{b,y} - I_s^{b,y-1} = k_s^{b,y} \quad (3.3)$$

$$\sum_y \sum_s k_s^{b,y} \leq 1 \qquad \forall b \in \Psi_b \quad (3.4)$$

### 3.2.2 Security Check Sub-problem in Electricity Network *(SP1)*

The solution of the master problem is sent to the security check sub-problem (SP1) shown in (3.5) -(3.25). In (3.5) the objective function captures the mismatches in real and reactive power between the demand and supply at each bus. The nodal real and reactive power balance between demand and generation considering the power consumption of data centers, i.e. $p_{n,sc}^{h,t,y}$, is enforced by (3.6) and (3.7), respectively. For the sake of simplicity, the reactive power consumption of data center is ignored considering the reactive power compensation assets exist in the data center facilities. In (3.8) and (3.9), the real and reactive power of the utility feeder are limited considering the allowable power factor at the feeder. In (3.10), the real power of photovoltaic units is limited to the forecasted values. In (3.11) the reactive power of photovoltaic units is restricted by the capacity of photovoltaic units' inverters. In (3.12) and (3.13), the real and reactive power of the distributed generation units are limited by their upper and lower bounds, respectively. In (3.14), (3.15) and (3.16) the real and reactive power as well as the available energy in the candidate BES system are limited to their respective capacities if they are selected to be installed. The stored energy in BES system is enforced to be within the acceptable range in (3.16) where $\eta$ is the ratio of the minimum to the maximum energy capacity of the BES system. By (3.17) and (3.18) the stored energy in BES system at the beginning and the end of operating period in each day is set to be equal to the $\sigma$ percent of the maximum energy of the storage unit. In (3.19) the charge/discharge of BES system at each hour is determined based on the amount of energy stored in or withdrawn from the battery during that hour. In (3.20) and (3.21) the real and reactive power flow in the lines under outage are forced to be zero. Constraints (3.22) -(3.25) pertain to simplified DistFlow model for radial distribution systems [3.23]. In (3.22) and (3.23), the difference of square values of voltage magnitude between the sending and receiving nodes of a line determines the real and reactive power flow of the line. Here, *M* is a large scalar. Constraint (3.24) addresses the limits of the bus voltages. In (3.25), the

apparent power flow through a line is limited to its maximum apparent power capacity. Constraint (3.25) is quadratic; therefore, the convexity set for this constraint is the area inside a circle with radius $S_l$ centered at the origin. Implementing the polygon-based linearization method, this circle is approximated by a convex polygon [3.24], with the radius $S_l^*$ calculated in (3.26), where $N$, is the number of sides for alternative generic polygon. Assuming $N = 6$, constraint (3.25) can be replaced with constraints (3.27) -(3.29).

$$\min W1 = \sum_{sc}\sum_{y}\sum_{b}\sum_{t}\sum_{h}\begin{pmatrix} Z^{+}{}_{b,sc}^{h,t,y} + Z^{-}{}_{b,sc}^{h,t,y} + \\ Z'^{+}{}_{b,sc}^{h,t,y} + Z'^{-}{}_{b,sc}^{h,t,y} \end{pmatrix} \tag{3.5}$$

$$s.t.$$

$$\sum_{i\in\gamma_i^b} p_{i,sc}^{h,t,y} + \sum_{pv\in\gamma_{pv}^b} p_{pv,sc}^{h,t,y} + \sum_{s\in\gamma_s^b} p_{s,sc}^{h,t,y} + \sum_{b\in\eta_b} p_{f,b,sc}^{h,t,y} - \sum_{d\in\gamma_d^b} p_{d,sc}^{h,t,y} - \sum_{n\in\gamma_n^b} p_{n,sc}^{h,t,y} + Z^{+}{}_{b,sc}^{h,t,y} - Z^{-}{}_{b,sc}^{h,t,y} = \sum_{l\in L_{f,b}} p_{l,sc}^{h,t,y} - \sum_{l\in L_{t,b}} p_{l,sc}^{h,t,y} \tag{3.6}$$

$$\sum_{i\in\gamma_i^b} q_{i,sc}^{h,t,y} + \sum_{pv\in\gamma_{pv}^b} q_{pv,sc}^{h,t,y} + \sum_{s\in\gamma_s^b} q_{s,sc}^{h,t,y} + \sum_{b\in\eta_b} q_{f,b,sc}^{h,t,y} - \sum_{d\in\gamma_d^b} q_{d,sc}^{h,t,y} + Z'^{+}{}_{b,sc}^{h,t,y} - Z'^{-}{}_{b,sc}^{h,t,y} = \sum_{l\in L_{f,b}} q_{l,sc}^{h,t,y} - \sum_{l\in L_{t,b}} q_{l,sc}^{h,t,y} \tag{3.7}$$

$$p_{f,b}^{min} \le p_{f,b,sc}^{h,t,y} \le p_{f,b}^{max} \quad \forall b \in \eta_b \tag{3.8}$$

$$-\tan(\phi_f)\cdot p_{f,b,sc}^{h,t,y} \le q_{f,b,sc}^{h,t,y} \le \tan(\phi_f)\cdot p_{f,b,sc}^{h,t,y} \ \forall b \in \eta_b \tag{3.9}$$

$$p_{pv,sc}^{h,t,y} \le \tilde{p}_{pv,sc}^{h,t,y} \tag{3.10}$$

$$q_{pv}^{min} \le q_{pv,sc}^{h,t,y} \le q_{pv}^{max} \tag{3.11}$$

$$p_i^{min} \le p_{i,sc}^{h,t,y} \le p_i^{max} \tag{3.12}$$

$$q_i^{min} \le q_{i,sc}^{h,t,y} \le q_i^{max} \tag{3.13}$$

$$-\hat{I}_s^{b,y}\cdot p_s^{max} \le p_{s,sc}^{h,t,y} \le \hat{I}_s^{b,y}\cdot p_s^{max} \quad : \overline{\tau p}_{s,sc}^{h,t,y}, \underline{\tau p}_{s,sc}^{h,t,y} \tag{3.14}$$

$$-\hat{I}_s^{b,y}\cdot q_s^{max} \le q_{s,sc}^{h,t,y} \le \hat{I}_s^{b,y}\cdot q_s^{max} \quad : \overline{\tau q}_{s,sc}^{h,t,y}, \underline{\tau q}_{s,sc}^{h,t,y} \tag{3.15}$$

$$\eta\cdot\hat{I}_s^{b,y}\cdot E_s^{max} \le E_{s,sc}^{h,t,y} \le \hat{I}_s^{b,y}\cdot E_s^{max} \ : \overline{\tau e}_{s,sc}^{h,t,y}, \underline{\tau e}_{s,sc}^{h,t,y} \tag{3.16}$$

$$E_{s,sc}^{1,t,y} = \hat{I}_s^{b,y}\cdot E_s^{max}\cdot\sigma \qquad : \tau_{s,sc}^{t,y} \tag{3.17}$$

$$E_{s,sc}^{1,t,y} = E_{s,sc}^{NH,t,y} \tag{3.18}$$

$$E_{s,sc}^{h,t,y} - E_{s,sc}^{h-1,t,y} = p_{s,sc}^{h,t,y} \tag{3.19}$$

$$-M.U_{l,sc}^{h,t,y} \le p_{l,sc}^{h,t,y} \le M.U_{l,sc}^{h,t,y} \tag{3.20}$$

$$-M.U_{l,sc}^{h,t,y} \le q_{l,sc}^{h,t,y} \le M.U_{l,sc}^{h,t,y} \tag{3.21}$$

$$V_{b,sc}^{h,t,y} - V_{a,sc}^{h,t,y} \le 2.\left(r_l^{b,a}.p_{l,sc}^{h,t,y} + x_l^{b,a}.q_{l,sc}^{h,t,y}\right) + M.\left(1 - U_{l,sc}^{h,t,y}\right) \qquad l \in L_{f,a}, l \in L_{t,b} \tag{3.22}$$

$$V_{b,sc}^{h,t,y} - V_{a,sc}^{h,t,y} \ge 2.\left(r_l^{b,a}.p_{l,sc}^{h,t,y} + x_l^{b,a}.q_{l,sc}^{h,t,y}\right) - M.\left(1 - U_{l,sc}^{h,t,y}\right) \qquad l \in L_{f,a}, l \in L_{t,b} \tag{3.23}$$

$$\underline{V_b} \le V_{b,sc}^{h,t,y} \le \overline{V}_b \tag{3.24}$$

$$\left(p_{l,sc}^{h,t,y}\right)^2 + \left(q_{l,sc}^{h,t,y}\right)^2 \le (S_l)^2 \tag{3.25}$$

$$S_l^* = S_l.\sqrt{\frac{2\pi}{N}/\sin(\frac{2\pi}{N})} \tag{3.26}$$

$$-\sqrt{3}.\left(S_l^* + p_{l,sc}^{h,t,y}\right) \le q_{l,sc}^{h,t,y} \le -\sqrt{3}.\left(p_{l,sc}^{h,t,y} - S_l^*\right) \tag{3.27}$$

$$-\frac{\sqrt{3}}{2}.S_l^* \le q_{l,sc}^{h,t,y} \le \frac{\sqrt{3}}{2}.S_l^* \tag{3.28}$$

$$\sqrt{3}.\left(p_{l,sc}^{h,t,y} - S_l^*\right) \le q_{l,sc}^{h,t,y} \le \sqrt{3}.\left(p_{l,sc}^{h,t,y} + S_l^*\right) \tag{3.29}$$

By solving SP1: (3.5) -(3.24), (3.27) -(3.29), if there is a mismatch in (3.5), i.e.,$\widehat{W}1 > 0$, an infeasibility cut will be generated and added to the master problem in (1)-(4). The infeasibility cut has the following form:

$$\widehat{W}1 + \sum_{sc}\sum_{y}\sum_{t}\sum_{s}\Big[\sum_{h}\Big((\overline{\tau p}_{s,sc}^{h,t,y} + \underline{\tau p}_{s,sc}^{h,t,y}) \cdot p_s^{max} \cdot \left(I_s^{b,y} - \hat{I}_s^{b,y}\right) + \left(\overline{\tau q}_{s,sc}^{h,t,y} + \underline{\tau q}_{s,sc}^{h,t,y}\right) \cdot q_s^{max} \cdot \left(I_s^{b,y} - \hat{I}_s^{b,y}\right) + \left(\overline{\tau e}_{s,sc}^{h,t,y} - \eta.\underline{\tau e}_{s,sc}^{h,t,y}\right) \cdot E_s^{max} \cdot \left(I_s^{b,y} - \hat{I}_s^{b,y}\right)\Big) + (\tau_{s,sc}^{t,y} \cdot \sigma \cdot E_s^{max} \cdot (I_s^{b,y} - \hat{I}_s^{b,y}))\Big] \le 0 \tag{3.30}$$

On the other hand, if a feasible solution is identified, the corresponding amount of supply dedicated for the data centers is passed to subproblem SP2 to verify its feasibility with respect to the data center operations. We describe this subproblem next.

### 3.2.3 Feasibility Check Sub-problem in Data Network *(SP2)*

This section presents the security check sub-problem of the data network, which is formulated in (3.31) -(3.34). The objective (3.31) is to minimize the mismatches between the total received requests from the customers at each period and the requests dispatched by the DCO through load balancer among the data

centers in the distribution network. In (3.32), non-negative slack variables (i.e. $Z^{\circ+}$ and $Z^{\circ-}$) measure the mismatch in the data network. Constraint (3.33) ensures that the workload assigned to each data center is less than its computing capacity. In (3.34), the electricity demand of data centers is determined based on the assigned workloads for each data center.

$$\min W2 = \sum_{sc}\sum_{y}\sum_{t}\sum_{h}\left(Z^{\circ+}{}_{sc}^{h,t,y} + Z^{\circ-}{}_{sc}^{h,t,y}\right) \tag{3.31}$$

$$s.t.$$

$$\sum_{n} \Gamma_{n,sc}^{h,t,y} + Z^{\circ+}{}_{sc}^{h,t,y} - Z^{\circ-}{}_{sc}^{h,t,y} = \Gamma_{tot}{}_{sc}^{h,t,y} \tag{3.32}$$

$$\Gamma_{n,sc}^{h,t,y} \leq \Gamma_{n}^{max} \tag{3.33}$$

$$\frac{\Gamma_{n,sc}^{h,t,y}}{\Gamma_{n}^{max}} \cdot p_{n}^{max} = \hat{p}_{n,sc}^{h,t,y} \qquad : \pi_{n,sc}^{h,t,y} \tag{3.34}$$

In case of any mismatch in (3.31), we compute an infeasibility cut that is added to subproblem SP1. This cut has the following form:

$$\widehat{W}2 + \sum_{n}\sum_{sc} \pi_{n,sc}^{h,t,y} . \frac{\Gamma_{n}^{max}}{p_{n}^{max}} . \left(p_{n,sc}^{h,t,y} - \hat{p}_{n,sc}^{h,t,y}\right) \leq 0 \tag{3.35}$$

The iterative process of solving SP1 to identify a feasible solution for distribution system operations continues until a feasible solution to the operations of data centers is identified. The feasibility of data center operations is indicted by slack variables taking zero values, or equivalently $\widehat{W2} = 0$.

### 3.2.4 Optimal Operation Sub-problem *(SP3)*

This sub-problem presents the optimal operation of the distribution system. Once the solution of the MP is feasible to both the electricity and the data networks, the feasible solution is passed to Optimal Operation Sub-problem (SP3). The objective for this sub-problem is to minimize the expected operational cost of distributed generation units and the expected cost of electricity supplied by the main distribution feeder, as shown in the first and second terms in (3.36), respectively. This sub-problem is subjected to constraints (3.8) -(3.24), (3.27) -(3.29) and (3.37), (3.38). Unlike (3.6) and (3.7) the power consumption for data centers are given in the nodal real and reactive power balance in (3.37) and (3.38).

$$\min W3 = \sum_{sc} \rho_{sc} \cdot \left( \sum_y \left(1 + d_y\right)^{1-y} \cdot \left[ \sum_t DT_t . \sum_h \left( \begin{matrix} \sum_i \left( \alpha_i \left( p_{i,sc}^{h,t,y} \right)^2 + \beta_i p_{i,sc}^{h,t,y} + \omega_i \right) \\ \sum_{b \in \eta_b} pr_{sc}^{h,t,y} . p_{f,b,sc}^{h,t,y} \end{matrix} \right) \right] \right) \quad (3.36)$$

$s.t.$

Constraints (3.8)- (3.24) and (3.27)- (3.29)

$$\sum_{i \in \gamma_i^b} p_{i,sc}^{h,t,y} + \sum_{pv \in \gamma_{pv}^b} p_{pv,sc}^{h,t,y} + \sum_{s \in \gamma_s^b} p_{s,sc}^{h,t,y} + \sum_{b \in \eta_b} p_{f,b,sc}^{h,t,y} {}_{in}^{h,t,y}{}_{b,sc} - \sum_{d \in \gamma_d^b} p_{d,sc}^{h,t,y} - \sum_{n \in \gamma_n^b} \hat{p}_{n,sc}^{h,t,y} =$$

$$\sum_{l \in L_{f,b}} p_{l,sc}^{h,t,y} - \sum_{l \in L_{t,b}} p_{l,sc}^{h,t,y} \quad (3.37)$$

$$\sum_{i \in \gamma_i^b} q_{i,sc}^{h,t,y} + \sum_{pv \in \gamma_{pv}^b} q_{pv,sc}^{h,t,y} + \sum_{s \in \gamma_s^b} q_{s,sc}^{h,t,y} + \sum_{b \in \eta_b} q_{f,b,sc}^{h,t,y} - \sum_{d \in \gamma_d^b} q_{d,sc}^{h,t,y} = \sum_{l \in L_{f,b}} q_{l,sc}^{h,t,y} - \sum_{l \in L_{t,b}} q_{l,sc}^{h,t,y} \quad (3.38)$$

Constraint (3.39) shows the optimality Benders cut, which is added to the master problem in (3.1) -(3.4) and updates the lower bound of the solution in the master problem. Here, $\widehat{W3}$ is the value of the objective function in (3.36).

$$Z_{lower} \geq \sum_{y=1}^{NY} \left(1 + d_y\right)^{1-y} . \left( \sum_s IC_s . k_s^{b,y} \right) + \widehat{W}3 + \sum_{sc} \sum_y \sum_s \sum_t \left[ \sum_h \left( (\overline{\tau p}_{s,sc}^{h,t,y} + \underline{\tau p}_{s,sc}^{h,t,y}) \cdot p_s^{max} \cdot \left( I_s^{b,y} - \hat{I}_s^{b,y} \right) + \left( \overline{\tau q}_{s,sc}^{h,t,y} + \underline{\tau q}_{s,sc}^{h,t,y} \right) \cdot q_s^{max} \cdot \left( I_s^{b,y} - \hat{I}_s^{b,y} \right) + \left( \overline{\tau e}_{s,sc}^{h,t,y} - \eta . \underline{\tau e}_{s,sc}^{h,t,y} \right) \cdot E_s^{max} \cdot \left( I_s^{b,y} - \hat{I}_s^{b,y} \right) \right) + \tau_{s,sc}^{t,y} \cdot \sigma \cdot E_s^{max} \cdot (I_s^{b,y} - \hat{I}_s^{b,y}) \right] \quad (3.39)$$

### 3.3 Solution Methodology

Figure 3-2 illustrates the flowchart of the solution to the proposed framework. The steps of the presented flowchart are as follows:

**Step a.** Solve *MP*, which provides an initial lower bound solution as $\hat{Z}_{lower}$ and set the upper bound solution $\hat{Z}_{upper}$ to an arbitrarily large scalar; go to step *b*.

**Step b.** Solve *SP1*, given the installation decisions $\hat{I}_s^{b,y}$; go to step *c*.

**Step c.** Solve the *SP2* given $\hat{p}_{n,sc}^{h,t,y}$ calculated in *SP1*; go to step *d*.

**Step d.** If $\widehat{W}2 = 0$, then go to step *f*. Otherwise, go to step *e*.

**Step e.** Add feasibility Benders cut (35) to the SP1 and go to step *b*.

**Step f.** If $\widehat{W}1 = 0$ then go to step *i*. Otherwise, go to step *g*.

**Step g.** Add feasibility Benders cut (30) to *MP*, go to step *h*.

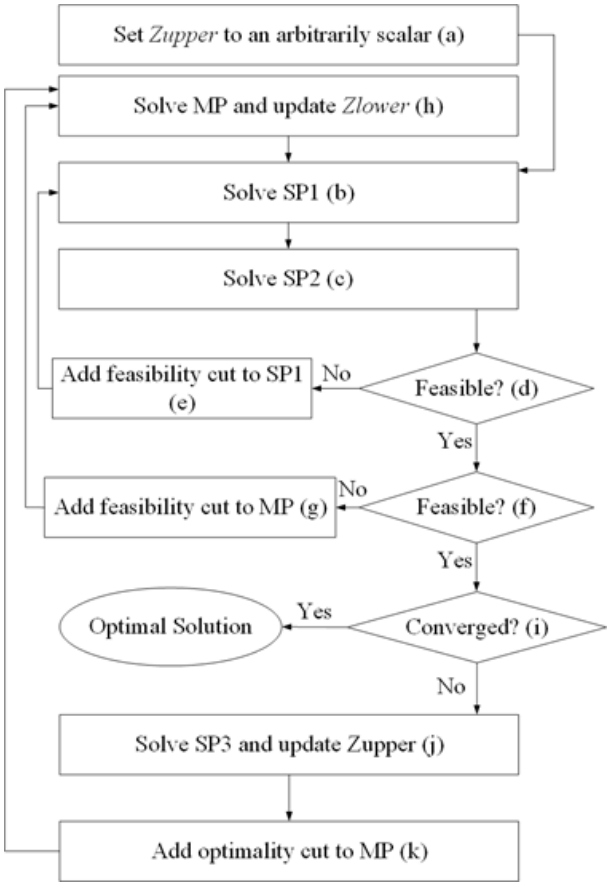

Figure 3-2 Flowchart for solving the proposed planning framework

**Step h.** Solve MP and update $\hat{Z}_{lower}$. Go to step *b*.

**Step i.** If $\left|\hat{Z}_{upper} - \hat{Z}_{lower}\right| \leq \epsilon$, then the optimal solution is procured, and the process stops. Otherwise, go to step *j*.

**Step j.** Solve *SP3* with given $\hat{p}_{n,sc}^{h,t,y}$ and $\hat{I}_s^{b,y}$. Update the upper bound solution as shown in (3.40), and go to step *k*.

$$\hat{Z}_{upper} = \sum_{y=1}^{NY}(1+d_y)^{1-y}.\left(\sum_s IC_s.\hat{k}_s^{b,y}\right) + \sum_{sc}\sum_y\sum_t\sum_s\left[\sum_h\left(\widehat{\overline{\tau p}}_{s,sc}^{h,t,y}\cdot\left(\hat{I}_s^{b,y}\cdot p_s^{max}\right) + \widehat{\underline{\tau p}}_{s,sc}^{h,t,y}\cdot\left(\hat{I}_s^{b,y}\cdot\right.\right.\right.$$

$$p_s^{max}\big) + \widehat{\widehat{\tau q}}_{s,sc}^{h,t,y} \cdot \big(\hat{I}_s^{b,y} \cdot q_s^{max}\big) + \underline{\widehat{\tau q}}_{s,sc}^{h,t,y} \cdot \big(\hat{I}_s^{b,y} \cdot q_s^{max}\big) + \widehat{\widehat{\tau e}}_{s,sc}^{h,t,y} \cdot \big(E_s^{max} \cdot \hat{I}_s^{b,y}\big) - \underline{\widehat{\tau e}}_{s,sc}^{h,t,y} \cdot \big(\eta \cdot E_s^{max} \cdot \hat{I}_s^{b,y}\big)\Big) + \hat{\tau}_{s,sc}^{t,y} \cdot \sigma \cdot E_s^{max} \cdot \hat{I}_s^{b,y}\Big] \quad (3.40)$$

**Step k.** Add optimality cut presented in (39) to the master problem and start the process from step *h*.

### 3.4 Case Study

In this section, an electricity distribution network with 20 buses and 19 lines shown in Figure 3-3 is used to evaluate the performance of the proposed planning framework. There are 3 distributed generation units, 2 photovoltaic generation units and 3 data centers in the distribution network. The capacities of photovoltaic units are 150 kW and 300 kW at buses B10 and B20, respectively. The capacities of data centers at buses 6, 13 and 18 are 100, 150 and 120 kW, respectively. Buses B2, B4, B10, B11, B13, B15 and B20 are considered as potential locations to install BES systems. The distribution network is connected to the main utility grid through bus B1. Tables 3-1, 3-2, and 3-3 present the characteristics of distributed generation units, BES systems, and distribution network's lines, respectively.

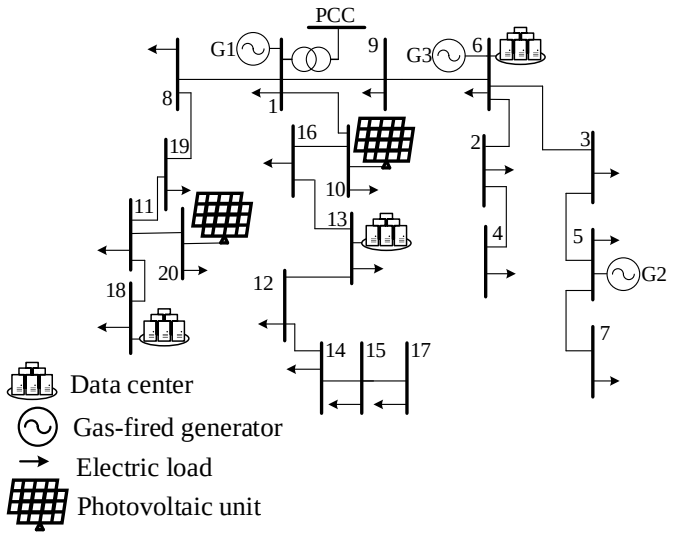


Figure 3-3 A 20-bus distribution network with data centers and photovoltaic units

Table 3-1 Distributed generation units' characteristics

| Unit | $P_{max}$ (kW) | $Q_{max}$ (kvar) | α ($/kW2h) | β ($/kWh) |
|---|---|---|---|---|
| G1 | 500 | 300 | 0.0001 | 0.004 |
| G2 | 300 | 200 | 0.0002 | 0.0333 |
| G3 | 100 | 70 | 0.0007 | 0.0075 |

Table 3-2 Battery energy storages' characteristics

| Battery | Pmax (kW) | Qmax (kvar) | Emax (kWh) | IC ($/kW) | $\sigma$ | $\eta$ |
|---|---|---|---|---|---|---|
| S1 | 200 | 120 | 400 | 9000 | 50% | 15% |
| S2 | 100 | 70 | 300 | 6000 | 50% | 15% |
| S3 | 50 | 30 | 200 | 2000 | 50% | 15% |

Table 3-3 Distribution network lines' characteristics

| l | From Bus | To Bus | Length (m) | Capacity (KVA) |
|---|---|---|---|---|
| 1 | 1 | 8 | 220 | 80 |
| 2 | 1 | 10 | 140 | 120 |
| 3 | 1 | 9 | 90 | 80 |
| 4 | 6 | 2 | 110 | 120 |
| 5 | 3 | 6 | 140 | 170 |
| 6 | 2 | 4 | 270 | 120 |
| 7 | 3 | 5 | 350 | 120 |
| 8 | 5 | 7 | 300 | 120 |
| 9 | 6 | 9 | 210 | 90 |
| 10 | 8 | 19 | 140 | 120 |
| 11 | 11 | 18 | 120 | 120 |
| 12 | 11 | 20 | 110 | 120 |
| 13 | 11 | 19 | 160 | 200 |
| 14 | 12 | 13 | 90 | 80 |
| 15 | 13 | 16 | 440 | 170 |
| 16 | 12 | 14 | 280 | 120 |
| 17 | 14 | 15 | 200 | 120 |
| 18 | 15 | 17 | 300 | 120 |
| 19 | 10 | 16 | 210 | 90 |

In the proposed planning framework, loads and price of electricity are used from PJM database [3.25]. The solar radiation data is imported from national solar radiation database (NSRDB) [3.26]. The trend for rate of arrived requests for each hour is simulated according to the requests made to the 1998 soccer World cup between April 30, 1998 and July 26, 1998 [3.27]. The proposed planning framework requires hourly loads, electricity prices, workloads of end users in data network, and output power of photovoltaic units. Here, Dissimilarity based Sparse Subset Selection (DS3) algorithm is leveraged to cluster the hourly data into a few subsets, that represent the original data based on the Euclidean distances as a guide of dissimilarity between the data points in data set [3.28]. In (41) matrix **R** represents the pairwise distances between the data points in data set as a measure of dissimilarities; given that there are *n* number of data points in the original data set.

$$\mathbf{R} = \begin{bmatrix} r_{1,1} & \cdots & r_{1,n} \\ \vdots & \ddots & \vdots \\ r_{n,1} & \cdots & r_{n,n} \end{bmatrix}_{n \times n} \tag{3.41}$$

The values of $r_{j,k}$ in **R** indicates how well the data points at row *j* represents the data point of column *k*; the smaller the value of $r_{j,k}$, the better data point in row *j* represents the data point at column *k*. In (3.42),

matrix **M** with the size of $n \times n$ is assigned to data set where $x_{j,k} \in [0,1]$ are variables that can be interpreted as the probability that data point at row *j* represents data point of column *k*. Here, $\boldsymbol{x_j^T}$ is the $j^{th}$ subset (i.e. representative**)** of matrix **M**.

$$\mathbf{M} = \begin{bmatrix} \boldsymbol{x_1^T} \\ \vdots \\ \boldsymbol{x_n^T} \end{bmatrix} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,n} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,n} \end{bmatrix}_{n \times n} \tag{3.42}$$

Mathematically, DS3 algorithm solves the minimization problem shown in (3.43) and (3.44). The objective of this optimization problem consists of two terms. The first term contributes to the number of the data subsets that should be selected, where $\|\boldsymbol{x_j}\|_l$ is the *l*-norm of the $j^{th}$ representative and $\mu$ is a tuning parameter, which affects the selected number of data subsets. Lower values of $\mu$ will lead to larger number of representatives. This translates to less error in representing the original data at cost of more data samples to deal with. The second term is the expected distance of the representatives from their corresponding data points in the original data set. Constraint (3.44) maps any data point of original data set to a limited number of representatives. Therefore, the solution of this optimization problem finds the least number of representatives that best represent the original data set.

$$\min_{\{x_{j,k}\}} \mu \cdot \sum_j^N \|\boldsymbol{x_j}\|_l + \sum_{k=1}^N \sum_{j=1}^N r_{j,k} \cdot x_{j,k} \tag{3.43}$$

$s.t.$

$$\sum_{j=1}^N x_{j,k} = 1\,,\ \forall k \tag{3.44}$$

The original data set for each year in our studies contains $4 \times 8760 = 35040$ data points including hourly loads, electricity prices, workloads of end users in data network, and output power of photovoltaics. Implementing DS3 algorithm this data set clustered into seven representative data subsets consisting of $7 \times 24 \times 4 = 672$ data points. Table 3-4 shows the representative days with the number of days represented in each year. Figure 3-4 shows the normalized output power of photovoltaic units for seven representative days utilized in the case studies.

Table 3-4 Time duration of reprehensive days

| Representatives | R1 | R2 | R3 | R4 | R5 | R6 | R7 |
|---|---|---|---|---|---|---|---|
| Hours | 1 | 6 | 14 | 39 | 71 | 112 | 122 |

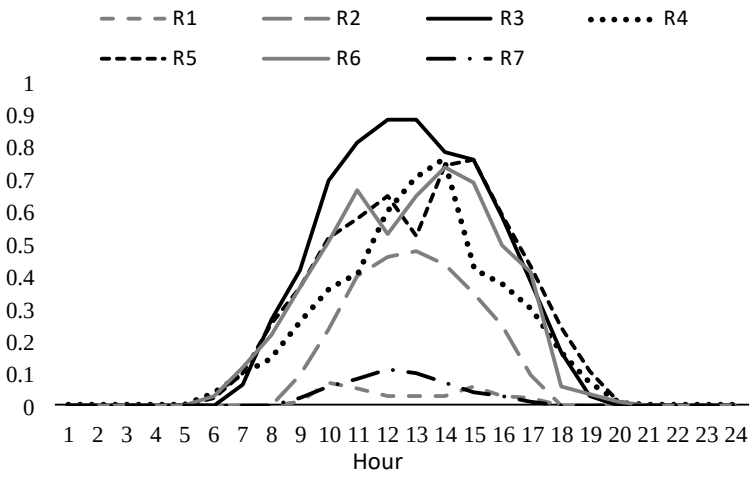


Figure 3-4 Normalized photovoltaic power outputs for representative days: R1-R7

The proposed expansion planning performed for 10 years, and

annual discount rate is 10%. The annual demand growth for data and electricity network is 5%. The following cases are considered:

Case 1 – Deterministic solution

Case 2 – Deterministic solution with contingencies

Case 3 – Stochastic solution

### 3.4.1 Case1- Deterministic Solution

In this case, no outage is considered in the distribution network. In the first iteration, the master problem is solved, and the solution is passed to SP1. In solution to SP1 yields the power supply for data centers i.e. $p_{n,sc}^{h,t,y} = 0$, which cannot serve the demand in the data network. Therefore, SP2 solved and the mismatch 3,720 indicates the deficiency in serving the demand in data network. The feasibility cuts (3.35) are generated and added to SP1. The number of iterations between SP1 and SP2 is 6. At the sixth iteration, the solution is feasible for the data network and the solution of the MP and SP2 leads to infeasibility in the electricity network. Therefore, feasibility cuts (3.30) are generated and added to the master problem. The total number of iterations between master problem and SP1 considering the solution to SP2 is 14. Once the solution is feasible for the electricity and data network, DSO's optimality check sub-problem is solved and the

optimality cuts are generated and sent back to the master problem. After one iteration between SP3 and master problem, the relative gap between the upper and lower bounds of the solution i.e. $\hat{Z}_{upper}, \hat{Z}_{lower}$ is less than the defined threshold $\epsilon = 0.005$ and the process stops. Table 3-5 shows the results of planning for this case. The total cost is $1,683,497. The installation cost of BES, the operation cost of distributed generation and the cost of purchased power from the main utility are $1,300,630, $270,253 and $112,614 respectively.

Table 3-5 Results of planning for Case 1

| Bus | Year | Storage type |
|---|---|---|
| B11 | 9 | S2 |
| B13 | 4 | S2 |
| B15 | 1 | S2 |
| B20 | 1 | S3 |

### 3.4.2 Case2- Deterministic solution with contingencies

In this case, the outages distribution lines L1, L2, L10, and L13 are considered. Table 3-6 shows the outages in the lines. The expansion plans for BES systems are shown in Table 3-7.

Table 3-6 Outages in Lines

| Line | Year | Representative | Hour |
|---|---|---|---|
| L1 | 4 | R3 | 13:00 |
| L2 | 8 | R1 | 18:00 |
| L10 | 5 | R5 | 14:00 |
| L13 | 9 | R3 | 16:00 |

Table 3-7 Results of planning for Case 2

| Bus | Year | Storage type |
|---|---|---|
| B10 | 9 | S2 |
| B11 | 1 | S3 |
| B13 | 5 | S2 |
| B15 | 1 | S2 |
| B20 | 4 | S3 |

Compared to Case 1, it is seen that one more storage unit installed at bus B10. The storage unit S2 at bus B11 replaced with smaller storage unit S3 but it is installed 8 years earlier and the storage unit at bus B20 installed 3 years later.

In Case 1 the loads on buses B10, B13 and B16 are supplied from the upstream network through line L2 and downstream network through line L4 at hour 18:00 of the first representative day in year 8. The load of data center at bus B13 is 56.3 kW in Case 1. At the same period in Case 2, line L2 is on outage, the loads on buses B10, B13, and B16 are supplied through the BES system S2 on bus B15 and the load of data

center decrease to 2.6 kW in Case 2. This indicates that the DCO dispatched the workloads to other data centers in Case 2 because of the outage in the electricity network at this period.

Figure 3-5a shows the power flow in the electricity network at hour 14:00 of fifth representative day in year 5 of planning in Case 1, whereas Figure 3-5b. shows the power flow for the same period in Case 2 when the line L10 is on outage. As seen in Figure 3-5a, the power of BES system S3 along with the generation of PV unit at bus B20 supply the upstream network. The power consumption of data center at bus B18 is 78.16 kW. As seen in Figure 3-5b, because of the outage in line L10, the power generation of PV unit is dedicated for charging the BES systems and the power consumption of data center reduced to 1.53 kW which indicates that DCO dispatches the workloads to other data centers.

The installation cost of BES increases from $1,300,630 in Case 1 to $1,331,676 in this case. Because of the increase in the total capacity of the BES systems in this case compared to Case1, the contribution of storage units to supply the demand in the electricity network is increased. As a result, the total operation cost of the electricity grid is decreased from $382,867 in Case 1 to $376,727 in this case. The operation cost of distributed generation and the cost of purchased power from the main utility is $263,417 and $113,310 respectively. The total cost for this case is $1,708,403 which is $24,906 higher than that in Case 1.

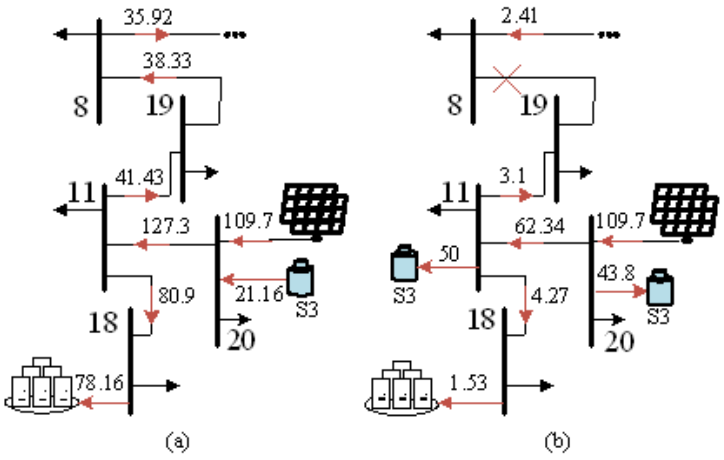


Figure 3-5 Flow of power in the electricity network. (a) Case1 with no outages. (b) Case 2 with outages

### 3.4.3 Case3- Stochastic Solution

This case represents the stochastic solution for the presented problem that captures the uncertainties in the electricity network's demand, electricity prices, the workload in the data network, generation of photovoltaic units. The forecast errors of distribution network demand, PV generation, the price of electricity, and the received workload by the DCO are captured by Gaussian distribution function. The mean values of the Gaussian distribution functions are equal to the forecasted values and the standard deviation is set to 3% of the mean values.

Using Monte-Carlo simulation, 3000 scenarios are generated, and scenario reduction techniques were used to reduce the number of scenarios to 10 effective scenarios with corresponded probabilities shown in table 3.8 [3.29], [3.30]. Table 3.9 shows the installation plan for BES systems considering the uncertainties. The installed capacity of BES system in Case 3 is 28% and 12.5% more compared with Case1 and Case 2, respectively. The installation cost of BES systems, the operation cost of distributed generation and the cost of purchased power from the main utility are $2,425,244, $292,870 and $105,968 respectively. The total cost for this case is $2,824,082 that is 40.3%, and 39.5% higher than total cost in Case 1, and Case2, respectively.

Table 3-8 Probability of scenarios

| Scenario | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| Probability (%) | 5.17 | 9.08 | 8.93 | 3.53 | 5.23 |
| Scenario | 6 | 7 | 8 | 9 | 10 |
| Probability (%) | 13.13 | 4.03 | 13.43 | 19.17 | 18.3 |

Table 3-9 Results of planning for Case 3

| Bus | Year | Storage type |
|---|---|---|
| 10 | 10 | S3 |
| 11 | 2 | S2 |
| 13 | 9 | S2 |
| 15 | 1 | S1 |

## 3.5 Summary

In this chapter, the expansion planning of BES systems in the distribution networks with data centers, and PV generation units is proposed. The proposed framework captures the interaction among the data center operator and distribution network operator to ensure the quality of service in the data network as well as the security of the distribution network. Benders decomposition is used to address the interaction between DCO

and DSO. The uncertainties in the demand in distribution network, electricity prices, workload in the data network, and the generation of PV units are considered. The effect of contingencies in distribution network, and the uncertainties in the planning horizon are addressed in the case studies.

### 3.6 References


[3.1] A. Shehabi, et. al. "United States Data Center Energy Usage Report," *Ernest Orlando Lawrence Berekeley National Laboratory*, June 2016.

[3.2] Asfandyar Qureshi, Rick Weber, Hari Balakrishnan, John Guttag, and Bruce Maggs. Cutting the Electric Bill for Internet-Scale Systems. *SIGCOMM Comput. Commun. Rev.*, 39:123–134, August 2009.

[3.3] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. H. Andrew, "Greening geographical load balancing," IEEE/ACM *Trans. Networking*, vol. 23, no. 2, pp. 657–671, Apr. 2015.

[3.4] K. Wang, M. Lin, F. Ciucu, A. Wierman, and C. Lin, "Characterizing the impact of the workload on the value of dynamic resizing in data centers," *SIGMETRICS Perform.* Eval. Rev., vol. 40, no. 1, pp. 405–406, Jun. 2012.

[3.5] W. Deng, F. Liu, H. Jin, C. Wu, and X. Liu, "MultiGreen: Cost minimizing multi-source datacenter power supply with online control," *in Proc. 4th Int. Conf. Future Energy Syst.*, Berkeley, CA, USA, 2013, pp. 149–160.

[3.6] A.Z. Amin "Renewable power generation costs in 2017," International Renewable Energy Agency (IRENA). *ISBN 978-92-9260-040-2, Copyright IRENA* 2018.

[3.7] A. Vafamehr, M.E. Khodayar, "Operation of distribution networks with volatile supply and controllable data center demand," IEEE/PES Transmission and Distribution Conference and Exposition (T&D),2018.

[3.8] J. Driesen and R. Belmans. "Distributed generation: challenges and possible solutions." 2006 *IEEE Power Engineering Society General Meeting (2006*).

[3.9] G. Carpinelli, G. Celli, S. Mocci, F. Mottola, F. Pilo and D. Proto, "Optimal Integration of Distributed Energy Storage Devices in Smart Grids," in *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 985-995, June 2013.

[3.10] Y. Zheng, Z. Y. Dong, F. J. Luo, K. Meng, J. Qiu and K. P. Wong, "Optimal Allocation of Energy Storage System for Risk Mitigation of DISCOs With High Renewable Penetrations," in *IEEE Transactions on Power Systems*, vol. 29, no. 1, pp. 212-220, Jan. 2014.

[3.11] M. Nick, R. Cherkaoui and M. Paolone, "Optimal Allocation of Dispersed Energy Storage Systems in Active Distribution Networks for Energy Balance and Grid Support," in *IEEE Transactions on Power Systems*, vol. 29, no. 5, pp. 2300-2310, Sept. 2014.

[3.12] M. Sedghi, A. Ahmadian and M. Aliakbar-Golkar, "Optimal Storage Planning in Active Distribution Network Considering Uncertainty of Wind Power Distributed Generation," in *IEEE Transactions on Power Systems*, vol. 31, no. 1, pp. 304-316, Jan. 2016.

[3.13] H. Saboori, and R Hemmati, " Optimal management and planning of storage systems based on particle swarm optimization technique, " *Journal of Renewable and Sustainable Energy*, vol. 8, no. 2, 2016.

[3.14] M. Daghi, M. Sedghi, A. Ahmadian, and M. Aliakbar-Golkar, "Factor analysis based optimal storage planning in active distribution network considering different battery technologies," *Applied Energy*, vol. 183, pp. 456-469, Dec 2016.

[3.15] S. F. Santos, D. Z. Fitiwi, M. Shafie-khah, A. W. Bizuayehu, C. M. P. Cabrita and J. P. S. Catalão, "New Multi-Stage and Stochastic Mathematical Model for Maximizing RES Hosting Capacity—Part II: Numerical Results," in *IEEE Transactions on Sustainable Energy*, vol. 8, no. 1, pp. 320-330, Jan. 2017.

[3.16] A. Pombo, J. Pina, V. Pires, "Multiobjective formulation of the integration of storage systems within distribution networks for improving reliability," *Electric Power system*, vol. 148, pp. 87-96, July 2017.

[3.17] J. Sardi, N. Mithulananthan, M. Gallegher and D. Quoc Hung, “Multiple community energy storage planning in distribution networks using a cost-benefit analysis” *Applied Energy* vol. 190, pp. 453-463, March 2017.

[3.18] A. S. A. Awad, T. H. M. EL-Fouly and M. M. A. Salama, "Optimal ESS Allocation for Load Management Application," in *IEEE Transactions on Power Systems*, vol. 30, no. 1, pp. 327-336, Jan. 2015.

[3.19] Q. Mingwen, et al, “Optimal planning and operation of energy storage systems in distribution networks with high wind power penetration,” *IET Center*, *Trans. Dist.* vol. 10, no. 8, 2019.

[3.20] I.v B, Yan W. “ Coordinated planning model of BESS and controllable switches in distribution, “ *Electronics letters*, vol. 50, no. 20, pp. 1479-1480, 2014.

[3.21] S. W. Alnaser and L. F. Ochoa, "Optimal Sizing and Control of Energy Storage in Wind Power-Rich Distribution Networks," in *IEEE Transactions on Power Systems*, vol. 31, no. 3, pp. 2004-2013, May 2016.

[3.22] Y. Zhang et al., “Optimal allocation of battery energy storage systems in distribution networks with high wind power penetration,” *IET Renewable Power Gener*., vol. 10, no. 8, pp. 1105–1113, Sep. 2016.

[3.23] M. E. Baran and F. F. Wu, "Network reconfiguration in distribution systems for loss reduction and load balancing," in *IEEE Transactions on Power Delivery*, vol. 4, no. 2, pp. 1401-1407, Apr 1989.

[3.24] H. Ahmadi and J. R. Marti, “Linear current flow equations with application to distribution systems reconfiguration,” IEEE *Trans. Power Syst.*, vol. 30, no. 4, pp. 2073–2080, Jul. 2015.

[3.25] Data Miner 2, PJM, [Online]. Available: https://dataminer2.pjm.com/feed/agg_definitions/definition

[3.26] NREL, NSRDB Data Viewer, available at: https://nsrdb.nrel.gov/nsrdb-viewer

[3.27] 1998 World Cup Web Site Access Logs - The Internet Traffic Archive, available at :http://ita.ee.lbl.gov/html/contrib/WorldCup.html

[3.28] E. Elhamifar, G. Sapiro, and S. S. Sastry, “Dissimilarity-based sparse subset selection,” arXiv preprint arXiv: 1407.6810 2014. Available at: http://arxiv.org/abs/1407.6810

[3.29] J. Dupacová, N. Gröwe-Kuska, and W. Römisch, “Scenario reduction in stochastic programming: An approach using probability metrics,” Math. Program. vol. 95, pp. 493–511, Jan. 2003.

[3.30] GAMS/SCENRED Documentation. Accessed: Aug. 2017. [Online].Available:https://www.gams.com/latest/docs/T_SCENRED.html

[3.31] A. Vafamehr, M. E. Khodayar and K. Abdelghany, "Oligopolistic Competition Among Cloud Providers in Electricity and Data Networks," in IEEE Transactions on Smart Grid, vol. 10, no. 2, pp. 1801-1812, March 2019.

Chapter 4

# ENERGY AWARE CLOUD COMPUTING

The structure of this chapter is as follows: Section 4.1 details the structure of cloud computing, including the various categories and layers of cloud computing, as well as different types of provided services. Section 4.2 surveys the impact of cloud computing on the quality and cost of the provided services indifferent sectors, the price of cloud computing services, and the notion of a cloud computing market. Section 4.3 explores energy management strategies at the macro level, including the interdependence of the cloud computing market and the wholesale electricity market. Energy management practices at the micro level, including the energy management practices in the data centers, are discussed in Section4.4. Section 4.5 is the summary.

## 4.1 Structure of Cloud Computing

The clouds are divided into four major models with distinct operational schemes: the public cloud, private cloud, hybrid cloud, and community cloud. In public clouds, the cloud computing resources are available to the public at the cost of lower security and privacy for the end users. Such models are not utilized by enterprises with high reliability and security requirements. In private clouds, the cloud computing resources are exclusively available to a single organization where the highest degree of control over performance, reliability, and security is offered to the end users at considerable operational cost. In this model, end users benefit from most features of cloud computing. In the hybrid cloud model, private IT resources dedicated to an enterprise are integrated with the public cloud. In this architecture, public and private infrastructure systems operate independently and communicate over the encrypted connections. This architecture enables the companies to store the protected data on private clouds while leveraging the computational resources in the public cloud to run applications that rely on the stored data. Such practices enhance the privacy and security of the data, as the data exposure to public domains is minimized. Furthermore, this architecture enables private corporations to leverage the resources of the public cloud once

their computing workloads exceed the computation capacity of the private cloud. Consequently, extra computing capacity could be acquired when such resources are needed. In community cloud architecture, the cloud computing services with their corresponding costs are mutually shared between several organizations with common apprehensions, privacy, and security requirements within a community. Example of such organizations include banks and trading firms such as NYSE Capital Markets Community; government organizations such as organizations in the State of California that share the computing infrastructure on the cloud to manage state residents' data; and the healthcare community cloud such as the QTS healthcare community cloud (HCC), which is designed to meet the unique needs of the healthcare industry. The structure of cloud computing is presented in the following four layers: (1) the hardware layer, (2) the infrastructure layer, (3) the platform layer, and (4) the application layer. Fig. 4-1 shows the layers of the cloud computing structure along with the cloud computing services provided at each layer.

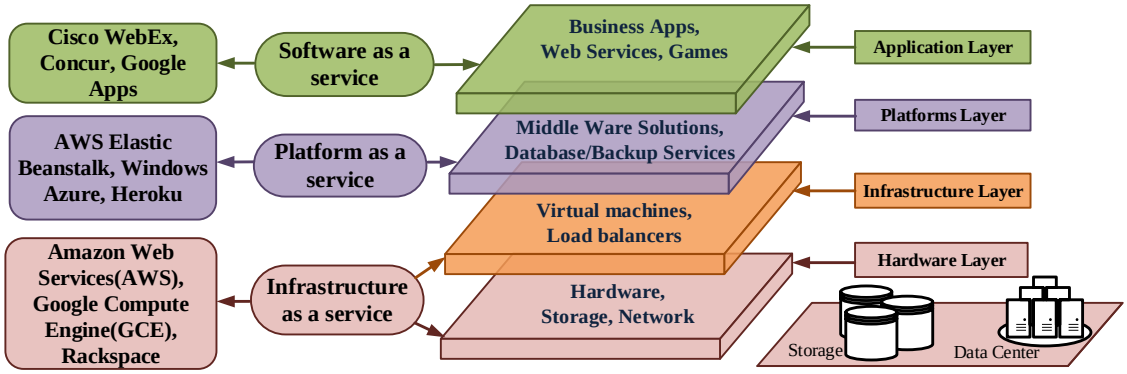


Figure 4-1 Structure of cloud computing

The application layer rests on the top of the cloud computing architecture. The application layer is the most visible layer for end users and contains the cloud's actual applications such as Google Apps, email, calendaring, and office tools (e.g. Microsoft Office 365). The platform layer is where the programming-language-level interface is supplied informs of operating systems and application frameworks to facilitate application deployment in cloud environments. The infrastructure layer is where virtualization technologies are employed to assign the required computing and storage resources dynamically. Examples of virtualization technologies include Xen [4.2], KVM [4.3], and VMware [4.4]. The hardware layer is established at data centers that consist of different physical components such as servers, routers, switches, as well as electricity

and cooling infrastructure. Such assets are managed and controlled by energy management systems. This architecture is perceived as a service-oriented model where each layer provides services to the above layer. In this perception, the services provided in cloud computing paradigm are categorized into three types: Infrastructure as a Service (IaaS), Plat-form as a Service (PaaS), and Software as a Service (SaaS). In IaaS, the computing resources, storage, and networking capabilities of service providers are offered to customers through a web-based graphical user interface. These services enable customers to leverage the required infrastructure for IT operation management. Amazon EC2, GoGrid, and Flexiscale are examples of IaaS providers. In PaaS, a platform is provided to allow customers to develop, execute, and manage applications without dealing with the complexities of building and maintaining the infrastructure associated with the development environments. Examples include Google App Engine and Microsoft Windows Azure. SaaS is a software licensing and delivery model in which the service providers host, manage, and offer the applications to customers over the internet at any time on a pay-for-use basis. SaaS eliminates the need for organizations to install and run applications on their own servers. Examples of SaaS providers include Rackspace, Hipchat, Trello, and Intercom.

### 4.2 Quality and Cost of Cloud Computing Services

In today's fast-paced world, different industries, businesses, and sectors require high standards of IT performance with an escalating burden of computing capabilities to enhance the quality of their services and products. As it is difficult and expensive for most businesses to build and maintain their own IT assets, they outsource such requirements to cloud providers that handle their considerable computing needs. The automotive, energy, manufacturing, education, entertainment, transportation, and healthcare sectors, as well as financial and insurance companies, are among the businesses that have embraced cloud computing for their day-to-day operation. Shop Direct, one of the UK's largest online retailers, adopts a hybrid cloud model to increase flexibility and reduce the response time to deliver more than 48 million products a year. Marriot International shifted toward cloud computing in today's internet era to meet the needs of travelers. Energy companies such as Shell Oil are using cloud services to manage and analyze the massive volume of geographical data granted from an installed base of sensors to detect and extract oil. Another important application of cloud computing is smart power grids. The smart grid is a combination of power networks

with intelligent entities that cooperate, communicate, and interact with each other to effectively deliver electricity to consumers. Achieving these objectives requires advanced technology for supporting mutual communication between the electric utilities and customers and processing real-time data for effective energy management strategies. Cloud computing provides the required computation asset for managing and processing the data collected from millions of smart meters in secure, reliable, and scalable manner. Considering the trending tendency toward cloud computing in various sectors, the re-liability, security, and economic efficiency of cloud computing have a considerable impact on the performance and objective measures in these sectors and consequently the social welfare. The vulnerability and risks associated with cloud computing services impact the quality and reliability of the cloud services that are provided for these sectors. As cloud technologies continue to mature, they share the same types of issues with in-house computing systems. However, any failure in the cloud computing services affects a larger population of users and has higher visibility compared to private in-house computing infrastructure systems. For instance, in November 2014 the Azure storage service was hit by a massive outage as a result of software updates for performance improvements. The outages in cloud computing services provided to the health and energy sectors can result in considerable economic loss and other severe consequences. The website HealthCare.gov that provides the health insurance frequently crashed when first launched due to flaws in the design and lack of resources to handle the excessive demand. On Oct. 21, 2016, Dyn, the world's most trusted in the DNS industry, suffered from a series of attacks and, as a result, dozens of websites and their businesses such as those of Airbnb, Twitter, Amazon, Ancestry, Netflix, and PayPal were affected. Such incidents can be triggered by a shortage in the energy supply or contingencies in the communication infrastructure systems, as the cloud computing infrastructure is heavily dependent on the reliability and security of these supporting systems. Furthermore, enterprises that use cloud services instead of local IT assets, have limited control over the quality, features, and cost of the provided services. Therefore, there is a considerable uncertainty in the cost, quality, and availability of the provided cloud services for such businesses that could lead to severe financial con-sequences for the clients that are heavily dependent on such cloud services. Another type of risk for adopting the cloud services is associated with the reliability of the internet connection. As cloud services are offered through the internet, any failure in the internet connections can cut the customers access to the cloud services. The DNS infrastructure is the weakness of the internet connectivity. There are 13 root

servers that act as a master list for the entire web's address book, and failure of any of these servers would lead to an extensive outage of the web domains. Another source of failures in the cloud computing infrastructure pertains to the data centers. Failures in air-conditioning and data center cooling systems can cause overheating and damages of the data storage devices. Such damages could affect financial transactions worth billions of dollars, traffic routing in a busy city, and emergency service call centers. Failures in various components of a power network inside a data center including uninterruptible power supplies (UPSs), power distribution units (PDUs), transformers, and breakers can lead to power outages. In June 2008, "The Planet" data-center in Houston hosting 9000 servers was down for several days as a result of an explosion and fire that were initiated by a short circuit in the power delivery network. In October 2009, a generator failed during planned maintenance at IBM's commercial data center in New Zealand, severely impacting Air New Zealand's check-in systems, online bookings, and call center systems that were outsourced to IBM. Google's datacenters in Belgium lost power in August 2015 after the local power grid was struck four times by lightning, causing outages for several days and a permanent loss of data for a small percentage of its end users. While most servers were supplied by redundant battery storage, 0.000001% of the disk space data was lost, which meant a few gigabytes for Google. Cloud providers handle various technical challenges in smart grids including optimizing energy management costs through monitoring and controlling the power grid assets, providing software applications on both the producer and consumer sides to control the power flow, and implementing various pricing strategies according to energy consumption, decreasing carbon emission by dispatching renewable energy re-sources effectively, and providing unlimited storage capacity for storing customers' data. The adoption of cloud services by power grid operators would result in more efficient and reliable delivering of electricity. However, the reliability and security of the provided grid services would be heavily dependent on the data and data processing cap-abilities of the cloud providers and any failure in the data centers that provide the cloud computing services can lead to considerable loss in the power grid. In 2003, the Northeast blackout was caused by a software flaw in an alarm system in a control room in Ohio that eventually led to a cascading failure. As a result, the energy supply was cut offto45 million people in eight states and 10 million people in Ontario. Cloud providers implement different pricing schemes for their offered services. The pricing strategy of each cloud provider is dependent on the strategies adopted by competitors. Offering higher prices for the same cloud service would result in losing the customers in the

cloud computing market. The pricing schemes for the cloud providers are divided into three categories: static, dynamic, and market-dependent. In the static pricing scheme, the customer pays a fixed price for the cloud services regardless of the volume of received services. In dynamic pricing scheme, the prices of cloud services vary dynamically with the service characteristics as well as customer characteristics and preferences. In market-dependent pricing schemes, the price of services is determined based on real-time market conditions including bargaining, auctioning, and demand behavior. Regardless of the choice of pricing scheme, the price of cloud computing services depends on several factors including the initial cost of the cloud resources, the quality of offered services including privacy and security, the availability of the resources, and operational and maintenance costs. Currently, the cloud computing market is an oligopoly among vendors such as Amazon, Microsoft, Google, and IBM. These cloud providers implement inflexible pricing schemes based on the duration of the service and usage threshold. The lack of standard application programming interfaces (APIs) for the provided cloud services restricts customers' choice. An API is a set of clearly defined methods, protocols, and tools devised for communication between various software components including routines, data structures, object classes, variables, and remote calls. Adopting a unified interface would lead to forming a market structure in which the cloud services are treated as commodities. SHARP, Tycoon, Bellagio, and Shirako are some examples of re-search projects that propose a unified market structure for cloud services.

### 4.3 Macro-level Energy Management Solutions For Cloud Service Providers

Cloud service providers, such as Google, Amazon, and Microsoft own and operate geographically dispersed data centers that ensure acceptable quality of service for end-users across the globe. The ability to reroute applications among multiple data centers is one of the important factors to provide secure, fast, and more available services to end users [4.5]. A geo-distributed cloud environment that runs over distributed data centers enables cloud providers to foster power management techniques with heterogeneous objectives. While the core idea for cloud computing management relies on geographically balancing the workload by redirecting the received requests from end users to the data centers, the decisions made on the volume of the workload being processed are dependent on the energy management objectives, including energy costs, carbon footprint, and sustainability. For instance, in order to mitigate the cost associated with power

consumption, the cloud computing workload can be redirected to datacenters located in colder areas with less cooling requirements [4.6]. To minimize energy costs, cloud providers leverage the spatiotemporal diversity of electricity prices in the wholesale market [4.7] and shift the data centers' workload to regions and periods with lower electricity prices. In order to produce the required electric energy for the data centers, hundreds of millions of tons of carbon dioxide are generated, meaning that attaining reductions in the carbon footprint of cloud providers is an important objective. For example, in order to supply energy to Google servers, it is estimated that 1.5 million tons of carbon dioxide were produced annually. For each query on the Google website, 0.2 g of carbon dioxide are emitted;10 min of viewing YouTube videos generates 1 g of carbon dioxide, a Gmail user on average produces 1.2 kg of carbon dioxide annually, and a typical user consumes 1.46 kg of carbon dioxide by using various Google services [4.8]. Redirecting the cloud computing workload to regions with abundant nuclear or renewable generation units would reduce the adverse environmental effects of cloud providers [4.9]. As the capacity of renewable generation is increasing in the United States, the intermittency and volatility in the energy supply is increasing. The Department of Energy plans to increase the current capacity of wind units from 5% to 20% of total U.S. electricity generation by 2030 [4.10]. While the generation of solar PV, biopower, geothermal, and hydropower was 11.6, 50,15, and 276 TWh in 2012, respectively, the energy produced from these sources is expected to increase to 234, 490, 184, and 431 TWh by 2030,respectively [4.11]. Data centers as large and controllable electric loads potentially can mitigate the intermittency of renewable generation. The data center load is flexible and shiftable in space and time and can be well coordinated with variable renewable generation in power networks. Despite the opportunities and benefits introduced by developing energy management schemes in data centers, the cloud workload management is a challenging task. Some cloud workloads impose data-intensive tasks for data centers that require operations on large datasets and use of special file systems such as the Hadoop distributed files system [4.12] and [4.13]. Leveraging such file systems would limit the dynamic shiftability of the workloads. The uncertainty in the received workloads further complicates the cloud workload management for the data centers and requires more complicated approaches to forecast the workloads. Cost-oriented workload scheduling and energy management may reduce the quality of service by increasing the latency in workload handling and by reducing the fault-tolerance. In order to maintain an acceptable quality of service and reliability, such criteria are considered as constraints in energy management practices. In order to implement

energy management strategies, cloud providers establish policies for workload distribution among the data centers. Energy management capabilities of the cloud providers promote demand response participation. The objective of the demand response programs is to dynamically manage the electricity demand of the consumers to keep the pace with supply side of the power system. Such programs can effectively reduce peak demand by 20% in the United States. Cloud providers can passively or actively participate in demand response programs. In passive participation, the data center demand is adjusted by leveraging various smart pricing approaches such as time-of-use pricing, peak pricing, and real-time pricing available in modern power systems. Cloud providers participate actively in demand response programs through bidding in different electricity markets (e.g. energy and ancillary service market), as well as proposing voluntary load reduction services to regional grid operators. Despite the opportunities for cloud providers to actively participate in demand response programs, there are several challenges including the immaturity of the markets and lack of appropriate regulations that limit their participation in the electricity markets. Participating in active demand response programs requires direct control load of grid opera-tors over data centers, which may degrade the performance of the services provided by cloud providers and result in financial losses. Consequently, complicated bidding and risk management strategies are required to motivate the cloud providers to participate in energy markets and mitigate their participation risks.

In large-scale data centers, energy-related costs are estimated to be 41.6% of the total operational cost, and the price of electricity impacts the cost of cloud services. Moreover, the price of electricity in the wholesale electricity market is affected by the substantial demand of the data centers themselves. While different pricing strategies and competition among cloud providers were addressed in the literature [4.14]-[4.16], the interdependence among the cloud computing market and the wholesale electricity market should be further highlighted. In [4.17], the global cloud market structure is introduced, where cloud providers offer their services into the cloud computing market, and end users submit their requests through brokers. The proposed market consists of a market directory that defines all the required cloud market rules, such as forms of payments, and workloads' required processing features such as the deadline for processing the workloads, and processing with or without interruption. The market directory receives and manages end users' requests and cloud providers' offerings in the cloud computing market. The auctioneers are responsible for clearing

the market and defining the share of the participants, and the banking system facilitates the transactions and market settlements. The interaction among the cloud computing and electricity markets is shown in Figure 4-2.

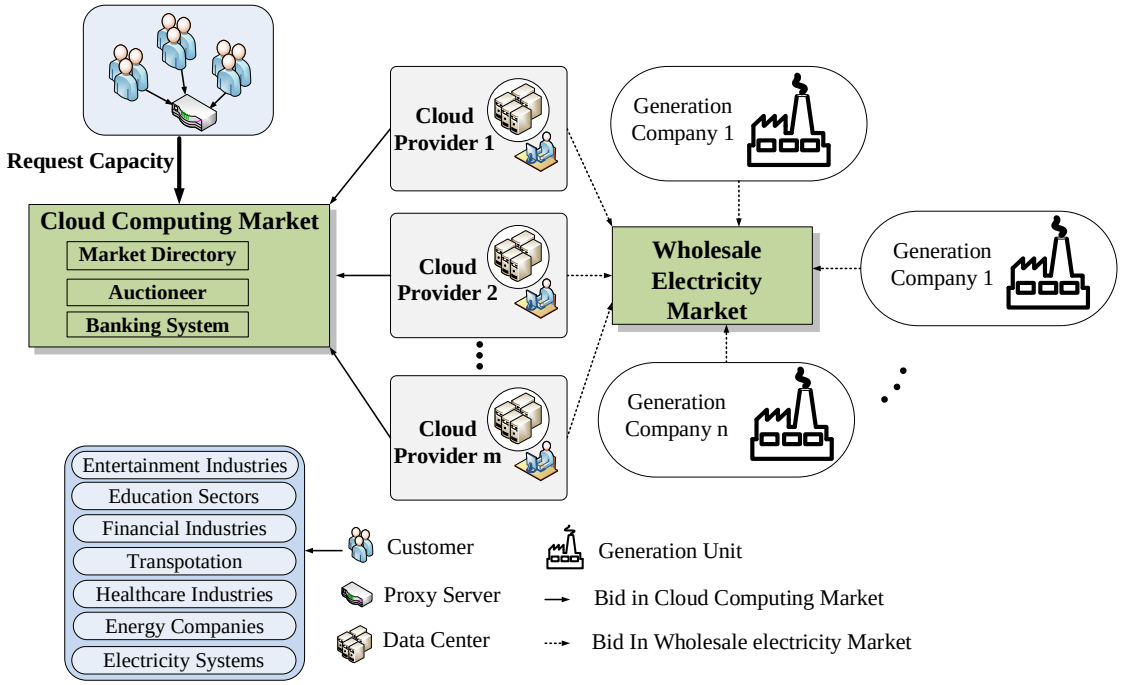


Figure 4-2 Cloud computing and wholesale electricity market structures

As shown in this figure, in cloud computing market, the cloud providers compete to offer their services to customers, while in the wholesale electricity market they bid for purchasing the required electricity to offer the cloud computing services. As the owners of data centers, the cloud providers lease their cloud resources to end users in the form of physical or virtual images of servers. The dynamic power consumption in the data centers corresponds to the CPU demand. Therefore, the volume of the processes performed in the data center servers would affect the required electric power for the cloud providers dramatically and managing these processes improves the economics and reliability of the provided services. The competition among the participants in each market can be perceived as a Cournot game. In the cloud computing market, the bidding strategy of the cloud providers could be determined by solving a bi-level optimization problem in which the upper-level problem maximizes their payoff while the lower-level problem captures the cloud computing market settlement process that maximizes the social welfare. In the

wholesale market, the objective of the cloud providers is to minimize their energy costs by bidding for their electricity demand. Similarly, the bidding strategy in the wholesale electricity market is determined by solving a bi-level optimization problem in which the upper-level problem maximizes the payoff while the lower-level problem captures the electricity market settlement performed by the in-dependent system operator (ISO). The ISO determines the price of electricity for the cloud providers by minimizing the operation cost or maximizing the social welfare. As the price of the cloud services de-pends on the price of electricity in the wholesale electricity market, the bidding strategy of the cloud providers in the wholesale electricity market affect the price of electricity and consequently the price of cloud computing services. Furthermore, the share of cloud providers and the price of cloud services are determined based on the bidding strategies of cloud providers in the cloud computing market. The share of the cloud provider in the cloud computing market determines the required electricity demand from the wholesale electricity market. Figure 4-3 represents the interactions among the cloud providers in the wholesale electricity and cloud computing markets.

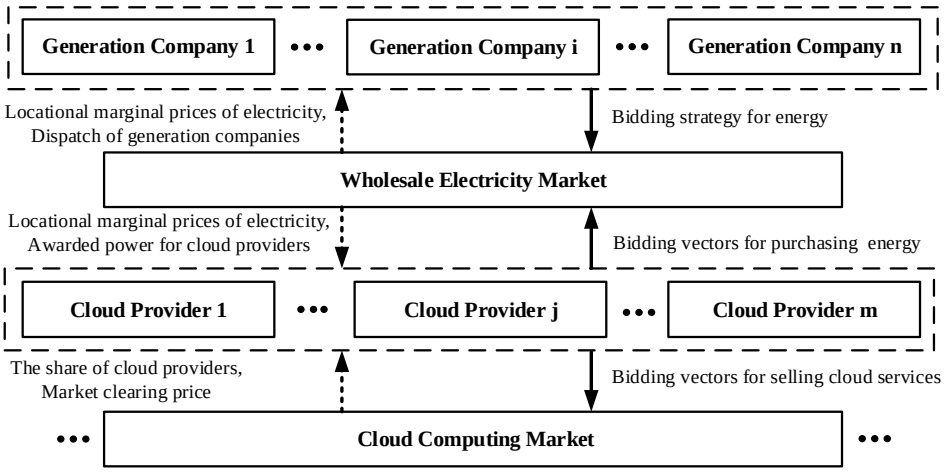


Figure 4-3 Strategic behavior of participants in the wholesale electricity and cloud computing markets

As shown in this figure, the cloud providers bid strategically for selling the cloud services in the cloud computing market, and buying the electricity from the wholesale electricity market. The bidding vector of the cloud providers in the cloud computing market captures the price of purchased electricity from the wholesale electricity market. The price of electricity is determined in the wholesale electricity market. As

bulk demand entities, the cloud providers also bid in the wholesale market to reduce their electricity costs. In this market, the generation companies along with cloud providers and other demand entities participate to maximize their payoffs while the ISO maximizes the social welfare. For each market participant, the bidding strategy could be determined by formulating a bi-level optimization problem. Implementing Karush-Kuhn-Tucker (KKT) optimality conditions for the linear lower-level problems transforms the bi-level optimization problems into single-level nonlinear optimization problems also known as mathematical problems with equilibrium constraints (MPEC). The MPECs with common lower-level problems for each market (i.e. the cloud computing market or wholesale electricity market) form equilibrium problems with equilibrium constraints (EPEC). The solution for the EPECs is the Nash equilibrium [4.18].

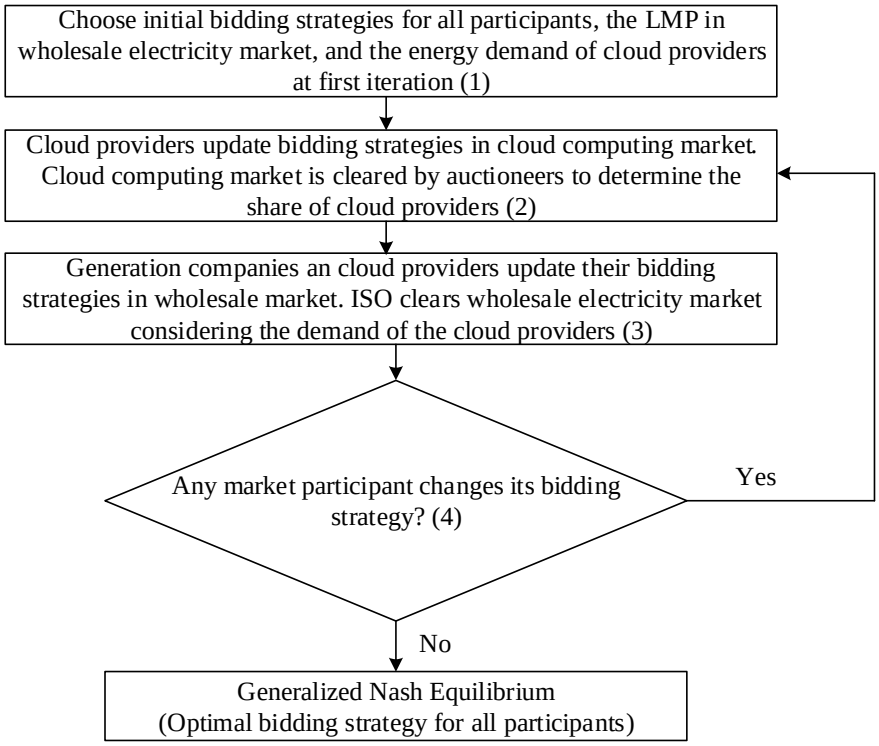


Figure 4-4 Solution framework to reach Nash equilibrium among the wholesale electricity and cloud computing markets

While an equilibrium could be achieved in the wholesale electricity and cloud computing markets, the interactions among these markets with entities participate in both, are captured as a dynamic game. The strategic behavior of the cloud providers in both markets highlights the interdependence among the electricity

and data network operations. Figure 4-4 represents the algorithm to solve for the Nash equilibrium in these interdependent markets.

The steps in the presented algorithm are as follows: In step (1), the bidding strategy of all participants in both markets, the LMP in the wholesale market and electricity demand of cloud providers are initiated. In step (2), the bidding strategy of the cloud providers is updated by calculating the sensitivity of their payoff functions with respect to the chosen strategies in the cloud computing market. The cloud computing market is cleared and the share of cloud providers, market cleared price in the cloud computing market, and upper bound of the electricity demand for cloud providers are determined. In step (3), knowing the upper bound of electricity demand for cloud providers calculated in step (2), the decisions made by generation companies and cloud providers are updated similarly based on the sensitivity functions of their payoffs in the wholesale electricity market. The wholesale electricity market is cleared and the price of electricity, the lower bound for the purchased electricity of cloud providers, and the share of the generation companies are determined. In step (4), if the sensitivity of the market participants' payoff with respect to the chosen strategies in both markets is zero, then the market participants would not change their bidding strategies and an equilibrium is reached. Otherwise, the algorithm continues in next iteration, starting at step (2).Attaining the equilibrium point entails solving two related EPEC problems in wholesale and cloud computing markets. In each EPEC, multiple nonconvex MPECs are solved. For this problem structure, there might be one solution, multiple solutions or no solution (i.e. Nash equilibrium), and the market regulations could impose additional bounds on the feasibility region of the presented problems that further facilitates the convergence and feasibility of the equilibrium [4.19], and [4.20].In the proposed solution framework in Figure 4-4, it is assumed the participants are rational players and each participant is aware of their opponents' payoff and cost functions. Under such assumptions, the problem is perceived as a complete information game. Practically in the wholesale electricity and cloud computing markets, the cloud providers and GENCOs have full knowledge of their own payoff and cost functions. However, partial information on the payoff and cost functions of the market participants is attainable through published data in both markets. Therefore, in practice, the problem is envisioned as an incomplete information game. Categorizing opponents' unknown strategies and characteristics into several "types" with respective realizations of payoff and cost functions,

and assigning a probability to each type will transform the incomplete information game into a complete game with imperfect information. In this paradigm, each participant maximizes the expected revenue considering different "types" of opponents with respective probability [4.21].Each cloud provider consists of multiple data centers distributed geographically to meet the workload of the end users, as shown in Figure 4-5. The requests received from end users are transferred to the load balancer, where the workloads are distributed among the data centers. The workload distribution scheme is dependent on the objective of the cloud providers to minimize the cost, maximize the redundancy and reliability, or minimize the carbon footprint. In order to determine the effective workload dispatch, data centers provide workload and energy signals to the cloud providers. The cloud provider determines the workload dispatch based on the signals received from the data centers to achieve the operational objectives.

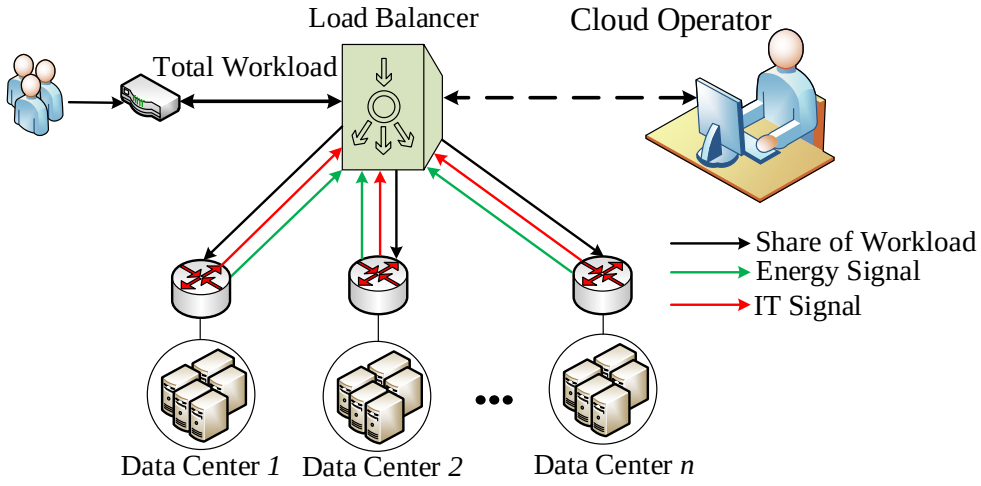


Figure 4-5 Energy and workload management among data centers

## 4.4 Micro-level Energy Management Strategies at Data Centers

Data centers are embedded in the hardware layer of the cloud computing architecture. A data center is composed of IT equipment including servers, storage devices, networking switches, and routers. The data network in a data center is built upon three basic layers known as core, aggregation, and access layers that are connected through switches and routers, as shown in Figure 4-6.

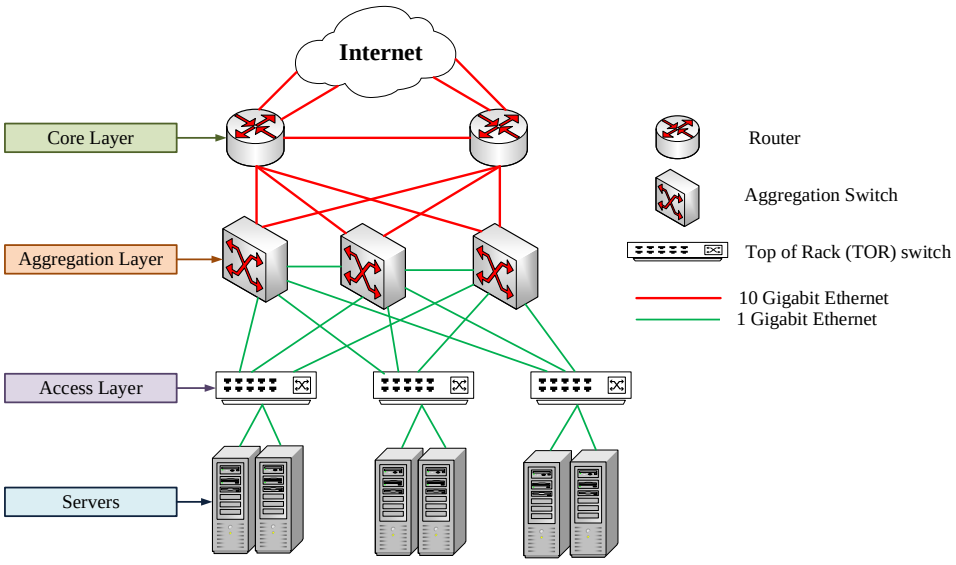


Figure 4-6 The network architecture in the data center

The core layer on top of the other layers serves as a gateway for connecting the data center to the internet and provides high-speed packet switching between the modules in the aggregation layer. The primary function of the aggregation layer is to aggregate thousands of packets entering or leaving the datacenter. This layer provides important tasks including domain and lo-cation services as well as server load balancing. At the access layer, the physical servers including web, application, and database servers are connected to the network. Besides the IT components, the data center is served by power cooling and humidification sub-systems, as shown in Figure 4-7. The power subsystem includes electrical switches, backup generators, renewable resources, transformers, UPSs, and PDUs. High-voltage AC power is transferred from the grid to the UPS through the transformer. UPSs are kept charged continuously to ensure the adequacy of energy supply for the IT components in short intervals after the failure of the grid supply and before starting the backup generator. From the UPS, the electricity is distributed among the PDUs at high voltage (480 V–1 kV).

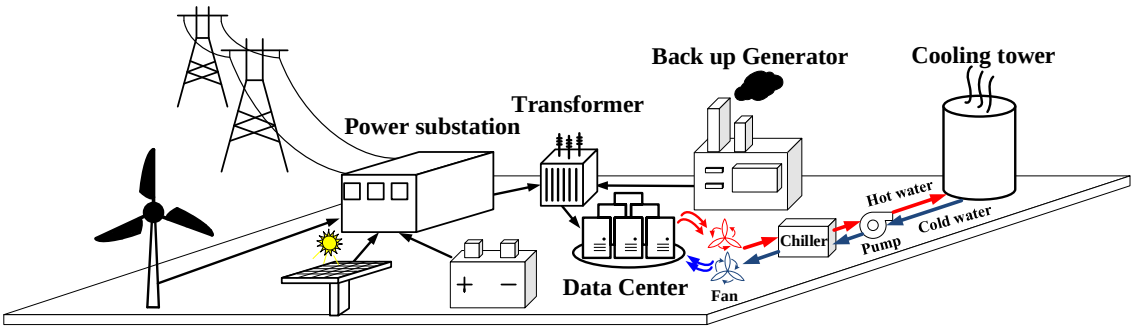


Figure 4-7 Energy infrastructure in the data center

The PDUs regulate the voltage to meet the IT equipment requirements. The energy loss in the data centers dissipates as heat and the generated heat is removed by the cooling infrastructure system. The cooling and humidification system consists of computer room air conditioners (CRACs), humidifiers, chiller plant, and cooling tower. CRACs absorb the hot air in the servers' room and blow the cold air. Humidifiers prevent electrostatic discharge in the servers and enable high-capacity, low-cost evaporative cooling. For the current data centers, the allowable range for relative humidity is 20–80% and a humidity level of 5.5 °C dew point to 60% relative humidity is recommended. Humidifiers are used in conjunction with free air cooling systems to increase the cooling capacity and to provide low-cost humidification of the air flowing through the ventilation system. Data centers consume a considerable volume of energy. The volume of energy consumed by an average-sized data center is equal to the energy consumed by 25,000 households [4.22]. The power consumption of data centers is approximately 50 times higher than the comparable office space. The energy consumption of the datacenters is a considerable portion of energy demand in the electricity network. In 2006, 1.5% of total electricity produced in the U.S. was utilized by data centers [4.23] and the power consumption of data centers is growing over 9% annually. The growth in electricity demand of data centers in the U.S. is 10 times larger than the overall growth of electricity demand. Shifting toward cloud computing and the emergence of new classes of applications such as satellite navigation and electronic shipment, electronic transactions in financial services, and high-performance scientific computing are among the reasons for such considerable growth rate. In the early stages of the structure development, performance indicators of the data centers were the only criteria for the cloud providers; however, the energy consumption and the corresponding expenses and carbon footprint that contribute to the total cost of ownership of the

cloud infrastructure systems, are becoming a predominant concern in recent years. Determining an effective model for energy consumption of a data center, as a function of the energy consumption of the sub-components and sub-systems, is not a straightforward process. The energy consumption of data centers depends on numerous factors including hardware con-figuration, cooling requirements, and types of applications running on servers. The energy consumed by the hardware, software, and the cooling and electricity infrastructure systems are all closely coupled. The energy consumption of a data center has two main parts: the energy consumption of IT assets such as servers, networks, and storage, and the energy usage by the cooling and power infrastructure systems. The volume of energy consumed by each sub-system varies with the design of the data centers and the energy efficiency of the subcomponents. Approximately, 60% of the total power consumption in a typical datacenter is associated with the IT hardware, including servers, storage devices, and network switches [4.24]; 30% of the total power consumption is related to the cooling infrastructure system, and8% of power consumption pertains to power loss in the data center power conditioning assets such as UPSs and PDUs. Several practices described in the following were established to improve the energy efficiency of the data centers.

### 4.4.1 Energy-Efficient IT Assets

The power consumption of servers has static and dynamic characteristics. The static power consumption captures the energy consumption of components such as memory, disk, and network interfaces, while the dynamic power consumption corresponds to the CPU power utilization that is further dependent on the operating voltage/frequency of the CPU and its utilization factor. CPUs have been regarded as the main contributors to the server's power consumption. Complementary metal-oxide semiconductor (CMOS) technology is a popular technology as it is resilient against noise and produces less heat during its operation compared to other semiconductor technologies. The power consumption of CMOS circuits used in microprocessors is in proportion to the frequency and the square of the operating voltage. Therefore, reducing the processor voltage would reduce the power consumption quadratically. However, reducing the supply voltage has negative impacts on the execution time of CPU. Reducing the operation voltage decreases the clock speed, which further increases the execution time of the workloads. Consequently, there is a trade off between the energy consumption and performance of the CPU using voltage scaling. The voltage and

frequency of CPUs are adjusted jointly using dynamic voltage and frequency scaling (DVFS) [4.25]. In the DVFS technique, the operation voltage and frequency of CPU are adjusted based on the intensity of workloads assigned to the servers. Leveraging DVFS would increase the efficiency of individual servers when they are not fully utilized. Another effective power management technique is dynamic cluster server configuration (DCSC), by which total power consumption of a cluster of servers is decreased. In this method, the workloads are consolidated on a subset of servers within a cluster and the rest are turned off to prevent the unused servers from wasting energy [4.26].The second largest power consumer in a server is the memory. The annual growth rate for storage devices is 17–24% and the rapid increase of memory's capacity leads to significant power consumption. Energy efficient approaches reduce data storage energy consumption by managing the storage capacity. Massive array of idle discs (MAID) is an effective power management solution as it saves power by shutting down idle discs and empowers the discs when an application needs to access data. Another tool to enhance the storage energy efficiency is automated storage provisioning. This storage management tool enhances the storage efficiency by detecting unused storage capacity and increasing the utilization of available storage devices. Another effective energy management solution is data compression. Data compression reduces the required capacity for storing data and further reduces the power consumption for storage. However, applying this method requires the data to be compressed before encryption on writes and decrypted before decompressing, which imposes energy overheads. Consequently, this storage management practice is recommended for files that are rarely accessed. It has been estimated that storage compression, consumes 15% to 30% less storage capacity [4.27]. Data deduplication is a fundamental practice in data storage that eliminates redundant data in a dataset. The volume of duplicated data in some enterprises reach 80% across the organization. By applying deduplication, extra copies of the same data are deleted, and by storing less data, fewer hardware resources consume energy. Another useful practice to reduce the energy consumption in memory is the storage snapshot technology. In this technique, snapshots taken from temporary copies of the data capture only the data changes instead of storing the complete copies of real-time data. Storage snapshots can save 80% to95% in energy use. Thin provisioning is another storage capacity management strategy that allocates the storage space centrally to the extent requested by the applications. Implementing this strategy will prevent over-providing of storage capacity and increases storage utilization from 30% to over 80%. This will reduce the required capacity of the storage devices by 40–60%.

Another storage management practice is to form tiered storage. In this method, data is stored on different disk types according to the relative demand for the data. Data with higher priority that is required to be accessed immediately will be stored in high-speed drives with considerable power consumption, and the low-priority data that is rarely accessed is stored on more energy-efficient storage devices with larger capacity and lower operating speed. Beyond the addressed methods for data storage management, energy-efficient storage hardware including lower speed and solid-state drives con-tribute to decreasing the energy consumption of the storage facilities. The higher spinning speed in high-performance disk drives such as 15 Krpm Serial Attached SCSI drives enables faster reading and writing at the cost of considerable power consumption. Using slower drives such as 7.5 K rpm Serial ATA drives reduces the energy consumption with negligible degradation in the performance. Implementing more expensive solid-state drives (SSDs) with no spinning discs improves the energy savings. Energy efficient servers leverage energy-efficient hardware including power supplies and fans, DC voltage regulators, and processors, as well as real-time measurement capabilities for power consumption, process utilization, and air temperature. These servers consume 30%less energy on average and deliver more processing power compared to the old technologies. Integrating more energy-efficient hardware will lead to less heat dissipation and consequently less energy consumption in the cooling infrastructure. For each watt saving in power consumption, one to two watts is saved in cooling power consumption. In the following, the energy efficiency in air conditioning and heat removal systems will be discussed.

### 4.4.2 Energy-Efficient Heat Removal and Air Conditioning

Cooling infrastructure is the second major source of energy consumption in data centers. Developing energy-saving strategies for the cooling systems provides considerable savings in energy consumption and cost. The number of servers within a rack ranges from 20 to 40. As data centers use highly dense server racks with 20–30 kW power demand per rack that generates 10 times more heat per square foot compared to older server racks, energy-efficient cooling systems play an important role in removing the dissipated heat with minimum energy consumption. Modifying the configuration of the servers' room to improve the airflow, utilizing visualization tools such as computational fluid dynamics models to optimizing the airflow, leveraging variable speed fans for cooling the servers, adjusting environmental set points to allow for wider

range of temperature and humidity within the range given by the server manufacturers are among the efforts to improve the energy savings in the cooling infrastructure system. The physical layer of the cooling infrastructure system consists of CRACs, pumps, cooling tower, and economizers in some cases, as shown in Figure 4-8.

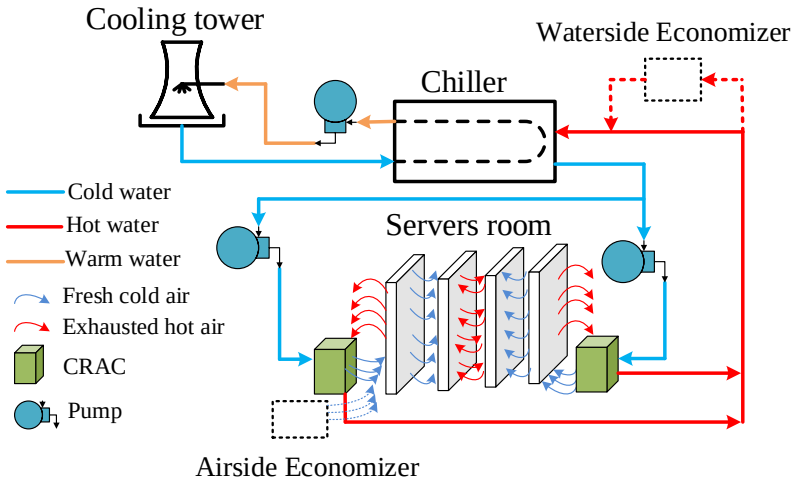


Figure 4-8 Cooling infrastructure of the data center

CRACs absorb heat from the servers' room and provide cold air. According to the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE), 10% of data centers' servers are too hot or too dry air-conditioned, and 40% of the cool air pumped into the servers' room is wasted [4.28]. Therefore, airflow management strategies contribute to energy savings in the cooling systems. Using variable-speed fans adjusts the airflow according to the configuration of the servers in the room and the loading of the servers. The air conditioning fans consume 5–10% of the total energy of the datacenter and 10% reduction in the speed of the fans would lead to 25%reduction in the fans' energy consumption. Forming hot and cold aisles in the server room improves the airflow management and reduces the energy loss. In order to do so, the placement of the server racks is modified such that the front and the back of the servers face each other, respectively. This configuration will prevent the air temperature in-crease through mixing the hot and cold air. Implementing physical barriers in forms of containment or rigid enclosures on server racks further mitigates the mix of cold and hot air. The containment systems can reduce the energy expenses by 5–10%. Adjusting diffusers to blow the cold air directly to the IT equipment, covering the unused space of racks

with blanking plates, leveraging structured cabling systems, in-stalling grommets on the floors where cold air is blown are among the efforts that enhance the recirculation of cold air through the server racks and improve the cooling efficiency. Temperature and humidity adjustment is an important factor in the energy management of the cooling infrastructure. The recommended temperature in data centers has been expanded to 65–80 °F, from 68 to 77 °F. However, many datacenters traditionally set their temperature as low as 55 °F. CRAC units use considerable energy to maintain the humidity; therefore, relaxing the acceptable humidity range would lead to considerable energy savings. Installing airside economizer enables exchanging the outside fresh air with inside hot air to reduce the burden on the cooling system. In a proof of concept presented by Intel IT, an airside economizer was utilized for cooling the servers by only using the outside air at temperatures of up to 90 °F and no significant difference between the failure rates of servers that were cooled using this technique compared to the servers that were cooled by the HVAC systems was reported. Datacenters in San Jose and Sacramento, Calif., reduced their cooling costs by 60% and 30% through airside economization, respectively [4.29]. Similarly, a waterside economizer that is integrated with chillers uses the evaporative capacity of the cooling tower to produce the chilled water. The economizer can substitute the chiller in winter to reduce the cooling cost by up to 70%.

### 4.4.3 Energy-Efficient Power Distribution

After IT assets and cooling infrastructure, power conditioning de-vices (e.g. UPSs and PDUs) are the main sources of energy loss. UPS improves the quality of the power during normal operation by mitigating the voltage sags, surges, and harmonic distortions. The main source of energy loss in UPS pertains to the transformer and the switching in the converters. In order to mitigate the switching losses, upgraded UPSs are equipped with a power management system that adjusts and optimizes the inverter's switching cycles considering the load level and type. Installing upgraded UPSs reduces energy losses by30–55%. For a high-efficiency 1000 kVA UPS, this means up to $18,000energy savings annually. The actual efficiency of a UPS depends on the power consumption of the IT assets supplied by that. The efficiency of UPS decreases dramatically in very light loads because of the fixed losses (e.g. losses in control logic units of the inverters). Based on an estimation offered by DOE, improving the efficiency of UPSs from 90%to 95% in a data center with a 15,000-square-foot area and 100-Wpower consumption per square foot would save $90,000 annually

at$0.12/kWh in energy bill [4.30]. PDU transfers reliable and conditioned power from UPS to the servers, networking equipment, and other IT facilities that need conditioned power. Maintaining higher voltage levels in UPS enables the designers to locate the PDU closer to the servers and electronic components and decrease of the conductor length and power loss. The high-voltage power delivered by UPS is converted to lower voltage in PDU by using a step-down transformer. Employing high-efficiency transformers at an optimized load factor is important to minimize the power loss as transformers operate most efficiently when they are loaded in 20%–50% of their nominal capacity and the efficiency slightly decreases as the loading exceeds 50%. Integrating input filters in conventional double conversion UPSs, mitigates the unwanted current harmonics and transformer loss [4.31].

## 4.5 Summary

This chapter highlights the interdependence among the electricity and cloud computing infrastructure systems by emphasizing energy-aware cloud computing. The challenges and opportunities for energy saving and energy management in the data centers are introduced. The energy management in data center enables the cloud providers to efficiently reduce the cost and improve the reliability measures of the cloud ser-vices. Cloud providers participate in the cloud computing market by offering the cloud services with respective offering prices. Similarly, they participate in the wholesale electricity market by offering bids for purchasing the required energy for the offered cloud services. The price of the purchased electricity impacts the offered price of cloud services in the cloud computing market. Furthermore, cloud computing demand itself impacts the required energy from the wholesale electricity market that would further impact the price of the purchased energy. To capture such interdependence, a hierarchical market structure is proposed where the cloud providers act as intermediate agents that participate in the wholesale electricity and cloud computing markets. In the proposed structure, cloud providers attempt to maximize their payoff in the cloud computing market while minimizing their costs in the wholesale electricity market by bidding strategically in both markets. The energy management strategies at the micro level at the data centers would help cloud providers to regulate electricity consumption, reduce cloud computing costs, and propose more competitive bids in the wholesale and cloud computing markets.

## 4.6 References


[4.1] P. Mell and T. Grance, “Final Version of NIST Cloud Computing Definition Published”, Oct. 2011. Available at: https://www.nist.gov/news-events/news/2011/10/final-version-nist-cloud-computing-definition-published

[4.2] XenSource Inc, Xen, www.xensource.com

[4.3] Kernal Based Virtual Machine, www.linux-kvm.org/page/MainPage

[4.4] VMWare ESX Server, www.vmware.com/products/esx

[4.5] Don MacVittie, “Maximizing the Strategic Point of Control in the Application Delivery Network,” F5 White Paper, 2012. Available at:http://be.security.westcon.com/documents/42226/f5-strategic-control-aplication-delivery.pdf

[4.6] J. Moore, J. Chase, P. Ranganathan, and R. Sharma, “Making Scheduling “Cool”: Temperature-Aware Resource Assignment in Data Centers,” *Proc. Usenix Ann. Technical Conf.,* Apr. 2005.

[4.7] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, “Cutting the electric bill for Internet-scale systems,” in Proc. ACM SIGCOMM, Barcelona, Spain, Aug. 2009, pp. 123–134.

[4.8] Duncan Clark. Google discloses carbon footprint for the first time. Available at: https://www.theguardian.com/environment/2011/sep/08/google-carbon-footprint.

[4.9] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. H. Andrew, “Greening geographical load balancing,” *in Proc. ACM SIGMETRICS*,2011, pp. 233–244.

[4.10] Department of Energy. 20% wind energy by 2030.
available at: https://www.nrel.gov/docs/fy08osti/41869.pdf

[4.11] IRENA REMAP 2030, a renewable energy roadmap, RENEWABLE ENERGY PROSPECTS: United States of America, available at: http://www.irena.org/REmap/IRENA_REmap_USA_report_2015.pdf

[4.12] A. Rahman, X. Liu, and F. Kong, "A Survey on Geographic Load Balancing Based Data Center Power Management in the Smart Grid Environment," *IEEE Comm. Surveys and Tutorials*, vol. 16, no. 1, pp. 214-233, first quarter 2013.

[4.13] D. Borthakur, The Hadoop Distributed File System: *Architecture and Design*, The Apache Software Foundation, 2007.

[4.14] V. Kantere, D. Dash, G. Francois, S. Kyriakopoulou, and A. Ailamaki, "Optimal Service Pricing for a Cloud Cache," *IEEE Trans. Knowledge and Data Eng.,* vol. 23, no. 9, pp. 1345-1358, Sept. 2011.

[4.15] M. Mihailescu and Y.M. Teo, "On Economic and Computational-Efficient Resource Pricing in Large Distributed Systems," *Proc. 10th Int'l Conf. Cluster, Cloud and Grid Computing (CCGrid),* pp. 838-843, 2010.

[4.16] Y. Feng, B. Li, and B. Li, "Price competition in an oligopoly market with multiple IaaS cloud providers," *IEEE Trans. Comput.,* vol. 63, no. 1, pp. 59–73, Jan. 2014.

[4.17] R. Buyya, C. ShinYeo, J. Broberg, and I. Brandic, "Cloud computing and emerging it platforms: Vision, hype, and reality for delivering computing as the 5th utility," *Future Generation Comput. Syst*., vol. 25, pp. 599–616, 2009.

[4.18] S. A. Gabriel, J. Zhuang, and R. Egging, "Solving stochastic complementarity problems in energy market modeling using scenario reduction," *Eur. J. Oper. Res.,* vol. 197, no. 3, pp. 1028–1040, 2009.

[4.19] H. D. Sherali, "A multiple leader Stackelberg model and analysis," *Oper. Res*., vol. 32, pp. 390–404, Mar. 1984.

[4.20] V. DeMiguel and H. Xu, "A stochastic multiple-leader Stackelberg model: Analysis, computation, application," *Manage. Sci.,* vol. 57, no.5, pp. 1220–1235, Sep. 2009.

[4.21] T. Li and M. Shahidehpour, "Strategic bidding of transmission constrained GENCOs with incomplete information," *IEEE Trans.Power Syst.,* vol. 20, no. 1, pp. 437–447, Feb. 2005.

[4.22] J. Kaplan, W. Forrest, and N. Kindler, "Revolutionizing Data Center Energy Efficiency," *Report by McKinsey&Company*, July 2008.

Available at: https://www.sallan.org/pdf-docs/McKinsey_Data_Center_Efficiency.pdf

[4.23] R. Brown, E. Masanet, B. Nordman, B. Tschudi, A. Shehabi, J. Stanley, J. Koomey, D. Sartor, P. Chan, J. Loper, S. Capana, B. Hedman, R. Duff, E. Haines, D. Sass, and A. Fanara, "Report to Congress on Server and Data Center Energy Efficiency," *Public Law*, vol. 109, p. 431, 2007.

[4.24] S. Pelley, D. Meisner, P. Zandevakili, T. F. Wenisch, and J. Underwood, "Power routing: Dynamic power provisioning in the data center," *ASPLOS*: *Architectural support for programming languages and operating systems*, 2010.

[4.25] J. S. Chase,D. C. Anderson, P.N. Thakar, A. Vahdat, and R. P. Doyle, "Managing energy and server resources in hosting centres," in *Proc. SOSP*, 2001.

[4.26] J. Li, Z. Li, K. Ren, and X. Liu, "Towards optimal electric demand management for internet data centers," *IEEE Trans. Smart Grid*, vol.3, no. 1, pp. 183–192, Mar. 2012.

[4.27] Energy Star, better management of data storage available at: https://www.energystar.gov/products/low_carbon_it_campaign/12_ways_save_energy_data_center/better_management_data_storage

[4.28] Robert F. Sullivan, Lars Strong P.E., and Kenneth G. Brill, "Reducing Bypass Airflow is Essential for Eliminating Computer Room Hot Spots.", Uptime Institute White Paper. available at: http://www.wilsonengineered.com/Tech/TUI811BypassAirflow_WP.pdf

[4.29] 42U, Solutions for the next generation Data Center, Air-side economizers available at: http://www.42u.com/cooling/air-side-economizers.htm

[4.30] Otto VanGeet, Best Practices Guide for Energy-Efficient Data Center Design, National Renewable Energy Laboratory (NREL), U.S. Department of Energy, 2011.

[4.31] Richard L. Sawyer, "Making Large UPS Systems More Efficient." *Elektron*, pp. 65-69, Jun. 2006.

[4.32] Pouria Razzaghi, Ehab Al Khatib, Shide Bakhtiari, Sliding mode and SDRE control laws on a tethered satellite system to de-orbit space debris, Advances in Space Research,2019.

Chapter 5

# OLIGOPOLISTIC COMPETITION AMONG CLOUD PROVIDERS IN ELECTRICITY AND DATA NETWORKS

This chapter proposes a framework to address the interaction among the cloud providers in the could computing market as well as the interactions among the cloud providers as demand entities in the wholesale market. The contributions of this chapter are as follows:

- The interaction among the electricity and cloud computing markets is captured using the presented oligopolistic model for competition and dynamic game in cloud computing and wholesale electricity markets.
- The strategic behavior of the market participants (i.e., cloud providers) in the cloud computing market; as well as the market participants (i.e., cloud providers and generation companies) in the wholesale electricity market to maximize their payoff were presented.
- A solution algorithm is proposed to determine Nash equilibrium for the proposed oligopolistic competition among the market participants in the wholesale electricity and could computing markets.

The presented formulation can be used to model the competition with incomplete information by associating respective probability measures to the types of the market participants. The rest of this chapter is organized as follows, Section 5.1 describes the structure and the interdependence among the wholesale and cloud computing markets. Section 5.2 presents the problem formulation for the multi-level dynamic game with complete information. The solution methodology is presented in Section 5.3 and the effectiveness of the proposed solution methodology is shown by a Case study in Section 5.4. The summary of this chapter is presented in Section 5.5.

## 5.1 List of Symbols

**Indices:**

| | |
|---|---|
| $b$ | Index of buses |
| $d$ | Index of electrical loads |
| $e$ | Index for the blocks on generation unit cost curve |
| $i$ | Index for generation companies (GENCOs) |
| $j$ | Index of cloud providers |
| $l$ | Index of power transmission lines |
| $m$ | Index of data centers |
| $n$ | Index of generation units |
| $s$ | Index for the blocks in power consumption curve of data center |

**Variables:**

| | |
|---|---|
| $P_{n,i}^{e}$ | Generation dispatch of segment $e$, generation unit $n$, GENCO $i$ [MW] |
| $P_{d}$ | Electricity demand [MW] |
| $P_{l}$ | Power flow in transmission line $l$ [MW] |
| $R_{i}$ | Payoff of GENCO $i$ [$] |
| $R_{j}$ | Payoff of cloud provider $j$ [$] |
| $f_{m,j}^{s}$ | Frequency of segment $s$ of data center $m$, cloud provider $j$ [THz] |
| $H_{(.)}$ | Power consumption function for data center |
| $P_{m,j}^{s}$ | Requested power at segment $s$ for data center $m$, cloud provider $j$ in CCM [MW] |
| $P'^{s}_{m,j}$ | Awarded power dispatch at segment $s$ for data center $m$, cloud provider $j$ in wholesale market [MW] |
| $k_{n,i}$ | Bidding strategy for unit n of GENCO $i$ |
| $k'_{m,j}$ | Bidding strategy for data center $m$, cloud provider $j$ in the CCM |
| $k"_{m,j}$ | Bidding strategy of data center $m$, cloud provider $j$ in the wholesale market |
| $\tau_{m,j}^{s}, \phi_{m,j}^{s}, \delta_{m,j}$ | Slack variables |
| $\mu_{n,i}^{e}$ | Bidding vector of segment $e$ for unit $n$ of GENCO $i$ [$/MWh] |
| $\mu'^{s}_{m,j}$ | Bidding vector of segment $s$ for data center $m$, cloud provider $j$ in the CCM [$/THzh] |

| | |
|---|---|
| $\mu''^{s}_{m,j}$ | Bidding vector of segment $s$ for data center $m$, cloud provider $j$ in the wholesale market [$/MWh] |
| $\rho_b$ | Locational marginal price at bus $b$ [$/MWh] |
| $\lambda_b, \lambda', \lambda_l^+, \lambda_l^-$ | Lagrange multipliers |
| $\theta^s_{m,j}, \omega_{m,j}$ | |

**Parameters:**

| | |
|---|---|
| $f_{req}$ | Total requested frequency by the end users [THz] |
| $\overline{f}^{s}_{m,j}$ | Capacity of segment $s$ for data center $m$, cloud provider $j$ [THz] |
| $\overline{k}_{n,i}, \underline{k}_{n,i}$ | Upper/lower limit of the bidding strategy for unit $n$ of GENCO $i$. |
| $\overline{k'}_{m,j}, \underline{k'}_{m,j}$ | Upper/lower limit of the bidding strategy for data center $m$, cloud provider $j$ in the CCM |
| $\overline{k''}_{m,j}, \underline{k''}_{m,j}$ | Upper/lower limit of the bidding strategy for data center $m$, cloud provider $j$ in the wholesale market |
| $\overline{P}^{e}_{n,i}, \underline{P}^{e}_{n,i}$ | Upper/lower limit of the generation unit cost curve segments [MW] |
| $SF^b_l$ | Shift factor of line $l$ with respect to injection at bus $b$ |
| $\psi^b_{(.)}$ | Set of components connected to bus b |
| $\beta'_{m,j}, \gamma'_{m,j}, \eta'_{m,j}$ | Coefficients for the power consumption function of data center $m$, cloud provider $j$ |

## 5.2 Market Structure

Figure 5-1 shows the structure of the cloud computing and wholesale markets. As shown in this figure, cloud providers participate in the wholesale market by bidding for their required energy. The demand and supply are balanced by the independent system operator (ISO) considering the security and reliability of the power network. The proposed structure captures the interdependence among the electricity and cloud computing markets. Figure 5-2 shows the proposed market structure for the cloud providers and GENCOs in the wholesale and cloud computing markets. GENCOs participate in the wholesale market to maximize their

profit, by bidding for the energy offered (μ) and cloud providers participate in both markets by strategically bid for cloud services ($\mu'$) and energy demand($\mu''$).

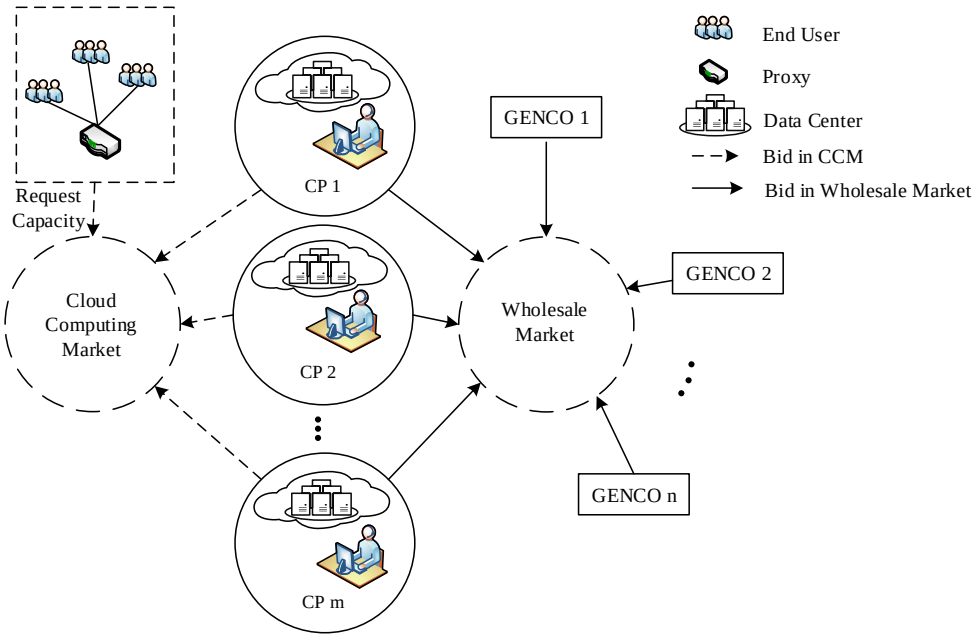


Figure 5-1 Market structure for electricity and cloud computing

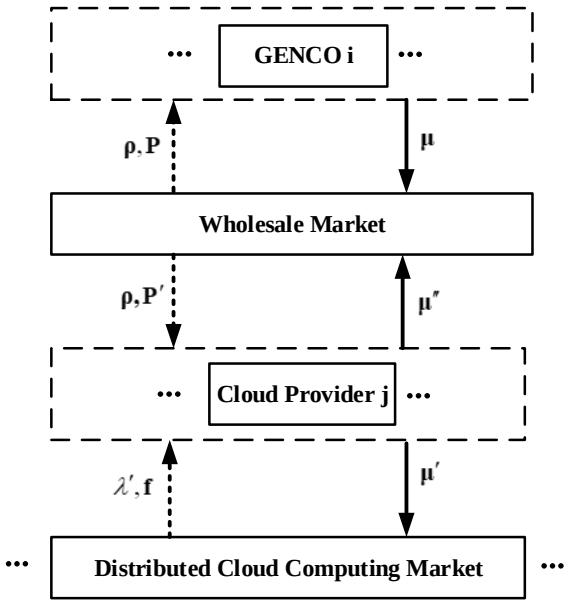


Figure 5-2 Bidding strategy and pricing in the hierarchical market structure

The CCM facilitates the balance among the demand and supply to determine the share of cloud providers ($\mathbf{f}$) and the market clearing price (MCP) for computing services ($\lambda'$). The cloud providers bid for the energy demand in the wholesale market to maximize their payoff while GENCOs compete for supplying energy. The social welfare is maximized in the wholesale market subjected to physical constraints of the market participants and the electricity network to procure the price of electricity, the GENCOs' dispatch, and the awarded dispatch for the cloud providers.

The link between the wholesale market and the CCM is the price of electricity and the electricity demand of the cloud providers. The price of services provided by the cloud providers is dependent on the price of electricity in the wholesale market. The price of electricity in the wholesale market is further affected by the bidding strategy of the cloud providers. The bidding strategy of the cloud providers for the provided cloud services in the CCM will affect their share as well as the price of these service in the CCM. The share of the cloud providers in the CCM will affect their electricity demand in the wholesale market. In order to determine the best strategy to maximize payoff, the market participants (GENCOs or cloud providers) are considered as rational players. Therefore, each market participant captures the behavior of the other participants assuming that they adopt strategies to maximize their payoff in the respective markets. The share of the cloud providers in the CCM is determined by the balance between end-users' demand and the supplied cloud services. The share of the cloud providers and GENCOs in the wholesale market is determined by the ISO. Here, each market participant has the knowledge of its own revenue and cost functions; therefore, the competition among the market participants in the wholesale market and CCM would yield Bayesian Nash equilibrium. Each market participant in the CCM and the wholesale market will categorize the unknown strategies and characteristics of other participants into defined "types". The market participants would represent the partial information on their opponents by associating probabilities for the opponents of being certain "type". In the wholesale market, GENCO *i*, and cloud provider *j*, have types $\mathbf{T_i}$ and $\mathbf{T_j}$ respectively. Hence, for GENCO *i*, $\mathbf{T}_{-i}^{G} = \mathbf{T}_1^G \times\cdots \mathbf{T}_{i-1}^G \times \mathbf{T}_{i+1}^G \times\cdots \mathbf{T}_{NG}^G \times \mathbf{T}_1^{CP} \times\cdots\times \mathbf{T}_{NCP}^{CP}$ denotes the type of GENCO *i*'s opponent strategy sets where *NG* and *NCP* are the total numbers of GENCOs and cloud providers respectively. Let $p(\mathbf{t}_{-i}|ti)$ be the conditional probability of the GENCO being of type $t_i$ and the opponents (i.e., other GENCOs and cloud providers) are being of type $\mathbf{t}_{-i}$. The vector of bidding strategy for GENCO

of type $t_i$, is $\mathbf{k}_i(t_i)$ and $\mathbf{k}_{-i}(\mathbf{t}_{-i})$ is the vector of the bidding strategy of GENCOs' opponents, including other GENCOs and cloud providers as their type is represented by $\mathbf{t}_{-i}$. This eventually transforms the non-cooperative competition with incomplete information into a non-cooperative competition with imperfect information formed by incorporating conditional probabilities associated with the set of realized "types" for the market participants. The solution for the incomplete information game among the market participants is similar to complete information game with the probabilities assigned to each set of the cloud provider and GENCO types. Hence, for GENCO $i$, the expected revenue for each type $t_i$ is shown as $\max \sum_{\mathbf{t}_{-i} \in \mathbf{T}_{-i}} Payoff_i(t_i, \mathbf{k}_i(t_i), \mathbf{k}_{-i}(\mathbf{t}_{-i})) \times p(\mathbf{t}_{-i}|ti) \times p(t_i)$. In this chapter, the information of the market participants in the CCM and wholesale market including the opponents' payoff and cost functions are known to all market participants and the presented formulation addresses the non-cooperative game with complete information which can be further extended to non-cooperative competitions with incomplete information as discussed above [5.26], [5.27].

### 5.3 Problem Formulation

In the wholesale market, GENCOs maximize their payoff, cloud providers minimize their payment, and ISO maximizes the social welfare considering the GENCOs', cloud providers' and power network's constraints. In order to determine the bidding strategy adopted by each market participant, a bi-level optimization problem is formulated in which the upper-level problem addresses the profit maximization while the lower level problem represents the social welfare maximization.

The lower level problem determines the LMPs and the supply and demand quantities in the wholesale market. In CCM, the cloud providers solve a bi-level optimization problem to determine their bidding strategy on the offered services. Here, the upper-level problem maximizes the participant's payoff while the lower level problem maximizes the net gain of all market participants by maximizing the social welfare.

The presented bi-level optimization problem is further transformed into a single-level problem, that is also referred to as a mathematical problem with equilibrium constraints (MPECs). The lower level problem is linear programming problem that can be represented by primal-dual constraints for the upper-level

problem. The non-cooperative oligopolies in the wholesale and cloud computing markets are represented as equilibrium problem with equilibrium constraints (EPEC) in which several MPECs with similar lowerlevel problems are solved. The solution for the proposed EPEC is a Nash equilibrium [5.28]. The EPECs are formulated for the wholesale electricity and cloud computing markets. Therefore, the Nash equilibrium for the presented market structure is the solution for two EPECs in the wholesale and cloud computing markets. In such structure, the bidding strategy of the market participants in one market will affect the bidding strategy of the market participants in the other market. In the following sections, the formulation to determine the bidding strategy of GENCOs, and cloud providers are presented.

### 5.3.1 Bidding Strategy of GENCOs in the Wholesale Market

The bi-level optimization problem for each GENCO is presented as (5.1)-(5.8), where (5.1)-(5.4) represent the upper-level problem. The objective function of the upper-level problem is the payoff of each GENCO as shown in (5.1). The payoff of the GENCO is the revenue (first term) minus the generation cost (second term). The variables in (5.1)-(5.4) are $P^{e}_{n,i}$ , $k_{n,i}$ and $\mu^{e}_{n,i}$. The limitations on the bidding strategy of generation units of GENCOs are determined in (5.2). In (5.3), the bidding vector for unit n of GENCO *i* is presented in which a piece-wise supply cost curve for the produced energy is used [5.29]. The dispatch at each segment is limited by a capacity limit as shown in (5.4).

$$R_i = \max_{P^{e}_{n,i},k_{n,i}} \left[\sum_{n\in N_i\cap\psi^{b}_{n}}\left(\sum_b \rho_b \cdot \sum_e P^{e}_{n,i} - \sum_e \alpha^{e}_{n,i}\cdot P^{e}_{n,i}\right)\right] \quad \forall i \tag{5.1}$$

*s.t.*

$$\underline{k}_{n,i} \le k_{n,i} \le \overline{k}_{n,i} \;\; \forall i,\forall n \tag{5.2}$$

$$\mu^{e}_{n,i} = k_{n,i}\cdot \alpha^{e}_{n,i} \;\; \forall i,\forall n,\forall e,\forall \mu^{e}_{n,i} \in \boldsymbol{\mu}_i \tag{5.3}$$

$$0 \le P^{e}_{n,i} \le \overline{P}^{e}_{n,i} \quad \forall i,\forall n,\forall e \tag{5.4}$$

$$\max \sum_j \sum_{m\in M_j} \sum_s \mu''^{s}_{m,j}\cdot P'^{s}_{m,j} - \sum_i \sum_{n\in N_i}\sum_e \mu^{e}_{n,i}\cdot P^{e}_{n,i} \tag{5.5}$$

$$\sum_i \sum_{n\in\psi^{b}_{n}}\sum_e P^{e}_{n,i} = \sum_{d\in\psi^{b}_{n}} P_d + \sum_j \sum_{m\in\psi^{b}_{m}}\sum_s P'^{s}_{m,j} \;\; ;\forall b \;:\lambda_b \tag{5.6}$$

$$\sum_s P'^{s}_{m,j} \geq \sum_s \underline{P}^{s}_{m,j} \quad \forall j, \forall m \quad P'^{s}_{m,j} \in \mathbf{P'}_j \tag{5.7}$$

$$P_l^{min} \leq \begin{pmatrix} \sum_b SF_l^b \cdot \sum_i \sum_{n \in \psi_n^b} \sum_e P_{n,i}^e \\ -\sum_b SF_l^b \cdot \sum_{d \in \psi_d^b} P_d \\ -\sum_b SF_l^b \cdot \sum_j \sum_{m \in \psi_m^b} \sum_s P'^{s}_{m,j} \end{pmatrix} \leq P_l^{max} \quad \forall l: \lambda_l^+, \lambda_l^- \; \forall m \tag{5.8}$$

$$\rho_b = \lambda_b + \sum_l SF_b^l \cdot (\lambda_l^+ - \lambda_l^-) \quad \forall b, SF_b^l \in \mathbf{SF}^{\mathrm{T}} \tag{5.9}$$

The lower level problem is shown in (5.5)-(5.8). Decision variables in (5.5)-(5.8) are $P_{n,i}^e$ and $P'^{s}_{m,j}$. The social welfare is maximized in (5). The first term in (5.5) is the cloud providers' willingness to pay (demand bid) in the wholesale market and the second term is the operation cost of the GENCOs with respective bidding vectors. The load balance constraint is shown in (5.6). Constraint (5.7) ensures that the cloud providers purchase sufficient energy from the wholesale market and therefore the awarded dispatch in the wholesale market is more than the energy required to provide computation services in the CCM. The capacity constraint of the transmission lines is presented in (5.8) where the shift factor (SF) is used to calculate the power flow in the transmission line [5.30]. The Lagrange multipliers of constraints (5.6) and (5.8) are applied to calculate the LMP at each bus $(\rho_b)$ as shown in (5.9).

### 5.3.2 Bidding Strategy of Cloud Providers in the CCM

In CCM, the cloud users rent "machines" from the cloud providers. A machine could be a physical server or a virtual image of a server [10]. The total energy consumption of the data center, which is operated by the cloud provider, includes the energy consumed by the IT equipment and the cooling system. The variable part of the energy consumption for the IT equipment is the energy consumed by CPUs. The power consumption of other components such as memory, disk, and the network interface is fixed [5.31]. The power consumption of the data center is presented as a quadratic function of provided frequency for the end users as shown in (5.10) where $\beta'_{m,j}, \gamma'_{m,j}, and\ \eta'_{m,j}$ are in MW/THz$^2$, MW/THz, and MW respectively [5.32]. The bi-level optimization problem for cloud providers in the CCM is presented as (5.11)-(5.18). Here, (5.11)-(5.14) are representing the upper-level problem and (5.15)-(5.18) are representing the lower level problem. The decision variables in (5.11)-(5.14) are $P_{m,j}^s$, $f_{m,j}^s$, $k'_{m,j}$ and $\mu'^{s}_{m,j}$. The payoff function of the cloud providers in CCM is described in (5.11), where the first term captures the revenue of each cloud provider and

the second term is the cost of power consumption to provide the required frequency. Here, the LMPs for energy are obtained from (5.9). The limits for bidding strategy and the bidding vector for each data center are presented in (5.12) and (5.13) respectively. The bidding vector of the cloud providers for the unit of the frequency offered in the CCM captures the price of electricity in the wholesale market. The quadratic frequency-electricity demand function for data centers in (5.10) is linearized using a piece-wise linear approximation as shown in (5.10) and (5.17).

$$P_{m,j} = H\left(f_{m.j}\right) = \beta'_{m,j} f^2_{m,j} + \gamma'_{m,j} f_{m,j} + \eta'_{m,j} = \sum_s \alpha'^s_{m,j} f^s_{m,j} \quad \forall j, \forall m \tag{5.10}$$

$$R_j = \max_{f_{m,j}, k'_{m,j}} \left[\sum_{m \in M_j} \lambda' \cdot \sum_s f^s_{m,j} - \sum_b \sum_{m \in \psi^b_m} \rho_b \cdot \sum_s P^s_{m,j}\right] \quad \forall j \tag{5.11}$$

*s.t.*

$$\underline{k'}_{m,j} \le k'_{m,j} \le \overline{k'}_{m,j} \;\; \forall j, \forall m \tag{5.12}$$

$$\mu'^s_{m,j} = k'_{m,j} \cdot \rho_b \cdot \alpha'^s_{m,j} \;\; \forall \mu'^s_{m,j} \in \boldsymbol{\mu}'_j, \forall j, \forall m \in \psi^b_m, \forall s \tag{5.13}$$

$$P^s_{m,j} \le \overline{P}^s_{m,j}, \quad P^s_{m,j} \in \mathbf{P}_j \quad \forall j, \forall m, \forall s \tag{5.14}$$

$$\min \sum_j \sum_{m \in M_j} \sum_s \mu'^s_{m,j} \cdot f^s_{m,j} \tag{5.15}$$

$$\sum_j \sum_{m \in M_j} \sum_s f^s_{m,j} = f_{req} \quad : \lambda' \tag{5.16}$$

$$f^s_{m,j} \le \overline{f}^s_{m,j} \quad ; \forall j, \forall m, \forall s \quad : \theta^s_{m,j} \tag{5.17}$$

$$\sum_s f^s_{m,j} \le \sum_s \frac{\overline{P}'^s_{m.j}}{\alpha'^s_{m,j}} \quad ; \forall j, \forall m \quad : \; \omega_{m,j} \tag{5.18}$$

The decision variables in (5.15) -(5.18) are $f^s_{m,j}$; and $\lambda'$, $\theta^s_{m,j}$ and $\omega_{m,j}$ are the Lagrange multipliers. In (5.15), the objective of the lower-level problem is to maximize the net gain of all participants by maximizing the social welfare in the CCM. In the proposed problem, the operation cost of the cloud providers is minimized. In (5.16) the total provided frequency by cloud providers in the CCM meets the end users' demand. The associated Lagrange multiplier of this constraint ($\lambda'$) shows the MCP of the provided services for the end users. The electricity demand-frequency function (5.10) is approximated using a piece-wise linear

function in which the frequency of each segment is limited as shown in (5.17). In order to provide the cloud services in the CCM, enough energy should be purchased from the wholesale market as enforced by (5.18).

### 5.3.3 Bidding Strategy of Cloud Providers in the Wholesale Market

Each cloud provider solves a bi-level optimization problem to minimize the energy payment. The upper-level problem is shown in (5.19) -(5.21) and the lower level problem is shown in (5.5)-(5.8). The decision variables in (5.19) -(5.21) are $P'^{s}_{m,j}, k"_{m,j}$ and $\mu"^{s}_{m,j}$. The energy payment is minimized as shown in (5.19) subjected to the bidding strategy constraints presented in (5.20). The bidding vector of each data center of each cloud provider to purchase electricity is shown in (5.21) where the cloud provider's bid in the wholesale market is a function of the price of cloud services in the CCM.

$$R_i = \min_{P'_{m,j}, k"_{m,j}} \left[ \sum_{m \in M_j \cap \psi_m^b} \sum_b \rho_b \cdot \sum_s P'^{s}_{m,j} \right] \quad \forall j \tag{5.19}$$

$$\underline{k"}_{m,j} \leq k"_{m,j} \leq \overline{k"}_{m,j} \quad \forall j, \forall m \tag{5.20}$$

$$\mu"^{s}_{m,j} = k"_{m,j} \cdot \lambda' \quad \forall j, \forall m, \forall s, \forall \mu"^{s}_{m,j} \in \boldsymbol{\mu}"_j \tag{5.21}$$

## 5.4 Solution Methodology

In the proposed market structure, cloud providers participate in the CCM and bid for the provided services to the end users. In the wholesale market, GENCOs compete to maximize their payoff by selling energy and the cloud providers compete for purchasing electricity to minimize their energy payments. The competition among the participants in each market is addressed as a Cournot game while interactions between the wholesale market and the CCM is presented as a dynamic game. To determine the bidding strategy of each cloud provider and procure the equilibrium, one requires to formulate problem as profit maximization subjected to middle-level and lower level problems that represent the market clearing problem for each market (i.e., wholesale market and CCM). In order to solve such three-level optimization problem, Karush-KuhnTucker (KKT) conditions associated with the market clearing problems should be formulated while ensuring the second order slater condition that requires the convexity of the formed KKT conditions. Formulating the complementarity conditions for one market in such three-level optimization problem would

result in forming a non-convex optimization problem. In other words, in a three-level optimization problem, if the bottom level's KKT (or other) optimality conditions are used to replace it in the middle-level problem, then the resulting middle-level problem is nonconvex due to the resulting complementarity conditions. To address this challenge, a solution framework is employed to determine the equilibrium point in the wholesale and cloud computing markets. The presented bi-level optimization problems are transformed into single level nonlinear optimization problems, by employing the KKT optimality conditions for the lower-level problems. The market participants update their bidding strategies by calculating the sensitivity of the payoff to the adopted bidding strategy in each market. Employing the KKT optimality conditions for the lower level problem (5.15)-(5.18), results in (5.22)-(5.29).

$$\lambda' + \tau_{m,j}^{s} - \theta_{m,j}^{s} - \omega_{m,j} = \mu_{m,j}^{\prime s} \qquad \forall j, \forall m, \forall s \tag{5.22}$$

$$\sum_j \sum_{m \in M_j} \sum_s f_{m,j}^{s} = f_{req} \tag{5.23}$$

$$f_{m,j}^{s} + \phi_{m,j}^{s} = \overline{f}_{m,j}^{s} \qquad \forall j, \forall m, \forall s \tag{5.24}$$

$$\sum_s f_{m,j}^{s} + \delta_{m,j} = \sum_s \frac{\overline{P}\prime_{m,j}^{s}}{\alpha_{m,j}^{\prime s}} \qquad \forall j, \forall m \tag{5.25}$$

$$f_{m,j}^{s} \cdot \tau_{m,j}^{s} = 0 \qquad \forall j, \forall m, \forall s \tag{5.26}$$

$$\phi_{m,j}^{s} \cdot \theta_{m,j}^{s} = 0 \qquad \forall j, \forall m, \forall s \tag{5.27}$$

$$\delta_{m,j} \cdot \omega_{m,j} = 0 \qquad \forall j, \forall m \tag{5.28}$$

$$\left(\tau_{m,j}^{s}, \theta_{m,j}^{s}, \phi_{m,j}^{s}, \delta_{m,j}, \omega_{m,j}, f_{m,j}^{s}\right) \geq 0 \qquad \forall j, \forall m, \forall s \tag{5.29}$$

$$\begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{I} & -\mathbf{I} & -\mathbf{I} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T_j^s} & \mathbf{F_j^s} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{\Theta_j^s} & \mathbf{0} & \mathbf{\Phi_j^s} & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{\omega}_\mathbf{j} & \mathbf{0} & \mathbf{0} & \boldsymbol{\delta}_\mathbf{j} \end{bmatrix} \begin{bmatrix} \frac{\partial \lambda\prime}{\partial k\prime_{m,j}} \\ \frac{\partial f_{m,j}^{s}}{\partial k\prime_{m,j}} \\ \frac{\partial \tau_{m,j}^{s}}{\partial k\prime_{m,j}} \\ \frac{\partial \theta_{m,j}^{s}}{\partial k\prime_{m,j}} \\ \frac{\partial \omega_{m,j}}{\partial k\prime_{m,j}} \end{bmatrix} = \begin{bmatrix} \frac{\partial \mu_{m,j}^{\prime s}}{\partial k\prime_{m,j}} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \tag{5.30}$$

Differentiating (5.22) -(5.29) with respect to the cloud provider's bidding strategies in the CCM will result in (5.30) where the coefficient matrix on the left-hand side is denoted by **M** and the right-hand side is represented by *d***y**.

Here, $\mathbf{F_j^s} = Diag(f_{m,j}^s)$, $\mathbf{\Phi_j^s} = Diag(\phi_{m,j}^s)$ , $\mathbf{T_j^s} = Diag(\tau_{m,j}^s)$, $\mathbf{\Theta_j^s} = Diag(\theta_{m,j}^s)$, $\boldsymbol{\omega_j} = Diag(\omega_{m,j})$, $\boldsymbol{\delta_j} = Diag(\delta_{m,j})$ and **I** is an identity matrix. For the system of nonlinear equations in (5.30), sensitivity function vector is denoted by: $d\mathbf{z} = \left[\frac{\partial \lambda'}{\partial k'_{m,j}}, \frac{\partial f_{m,j}^s}{\partial k'_{m,j}}, \frac{\partial \tau_{m,j}^s}{\partial k'_{m,j}}, \frac{\partial \theta_{m,j}^s}{\partial k'_{m,j}}, \frac{\partial \omega_{m,j}}{\partial k'_{m,j}}\right]^T$ and calculated as:

$$d\boldsymbol{z} = \mathbf{M}^T(\mathbf{M}\mathbf{M}^T)d\boldsymbol{y} \tag{5.31}$$

Exploiting the sensitivity functions obtained in (5.31), the payoff of the cloud provider is maximized in an iterative approach. In (5.32) the sensitivity of the cloud provider's payoff with respect to the chosen bidding strategy in the CCM is calculated. The components of (5.32) are calculated in (5.33) and (5.34).

$$\frac{\partial R_j}{\partial k'_{m,j}} = \frac{\partial R_j}{\partial f_{m,j}^s}\frac{\partial f_{m,j}^s}{\partial k'_{m,j}} + \frac{\partial R_j}{\partial \lambda'}\frac{\partial \lambda'}{\partial k'_{m,j}} \qquad \forall j, \forall m, \forall s \tag{5.32}$$

$$\frac{\partial R_j}{\partial f_{m,j}^s} = \lambda' - \rho_b \cdot \alpha_{m,j}'^s \;\; \forall j, \forall s, \forall m \in \psi_m^b \tag{5.33}$$

$$\frac{\partial R_j}{\partial \lambda'} = \sum_s f_{m,j}^s \quad \forall j, \forall m \tag{5.34}$$

Similarly, employing the KKT conditions for the lower level problem (5.5)-(5.8) representing the wholesale market and incorporating them in the upper-level problems (5.1)-(5.4) and (5.19)-(5.21) will transform these bi-level optimization problems to single level nonlinear programming problems. Utilizing similar approach, the sensitivity of the cloud providers' and GENCOs' payoffs with respect to the chosen bidding strategies in the wholesale market are calculated. Figure 5-3 shows the flowchart of the presented solution framework for this problem. The proposed structure is a dynamic game as cloud providers observe the price of electricity in the wholesale market and update their decisions in the CCM [5.33]. Similarly, the cloud providers observe the MCP and their share for could services in the CCM and update their bidding strategy in the wholesale market. A binary parameter $\zeta$ as observability indicator is introduced that captures the state of the cloud providers in one market once the price of the commodities in the other market is observed at each iteration. The steps of the presented flowchart are described as follows.

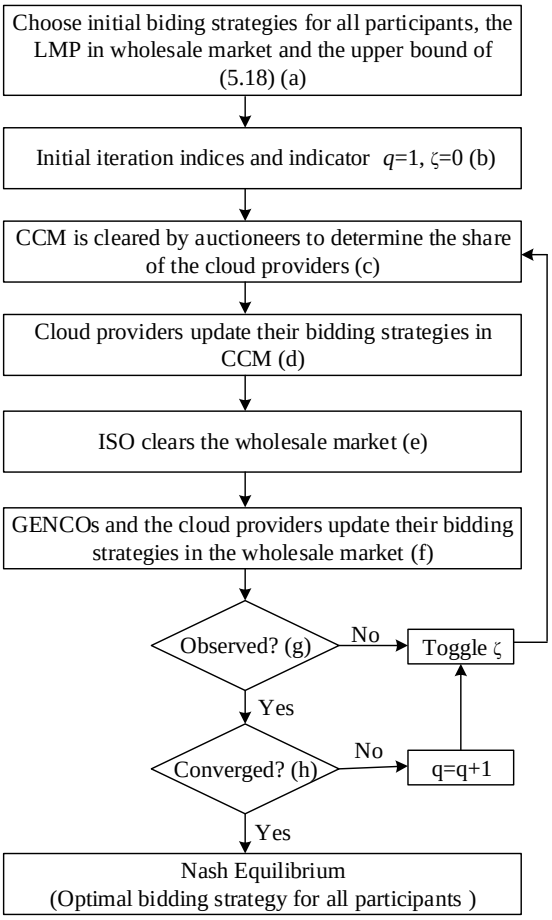


Figure 5-3 The solution framework for dynamic game in the wholesale market and CCM

In step (a), the bidding strategy for all participants in both markets, i.e., $k, k'$, and $k$", the LMP in the wholesale market, and the upper bound of (5.18) (i.e., $\overline{P'}^{s}_{m.j}$ ) are initiated.

In step (b), the iteration index $(q)$ and the observability indicator of the cloud providers $(\zeta)$ are initiated as $q = 1$ and $\zeta = 0$.

In step (c), the LMPs are set as (5.35) and the power capacity to supply the cloud services is determined by (5.36). The CCM is settled by solving (5.15)-(5.18) to determine the share of each cloud provider $\left(f^{s}_{m,j}\right)^{q}$, the MCP of cloud services $\lambda'^{q}$ in the CCM and the energy required to provide cloud services $\left(P^{s}_{m,j}\right)^{q}$. Go to step (d).

$$\rho_b^q = \rho_b^{q-1+\zeta} \qquad \forall b \tag{5.35}$$

$$\left(\bar{P}'^{s}_{m,j}\right)^{q} = \left(P'^{s}_{m,j}\right)^{q-1+\zeta} \quad \forall j, \forall m, \forall s \tag{5.36}$$

In step (d), the bidding strategy of the cloud providers $(k')$ in the CCM is updated using (5.37) by employing the sensitivity functions of their payoff with respect to their bidding strategy shown in (5.32).

$$k'^{q+1}_{m,j} = k'^{q}_{m,j} + \zeta \cdot \sigma \cdot \frac{\partial R^{q}_{j}}{\partial k'^{q}_{m,j}} \quad \forall j, \forall m \tag{5.37}$$

Here, $\sigma$ is a scalar that determines the step size for updating the bidding strategy. The bidding strategy of the cloud providers will not be updated prior to the realization of $\rho^{q}_{b}$ and $\bar{P}'^{s}_{m,j}$from the wholesale market. Go to step (e).

In step (e), the Lagrang multiplier $\lambda'$ and the lower bound for the purchased electricity are updated as shown in (5.38) and (5.39) respectively. The wholesale market is cleared by solving (5.5)-(5.8) using the updated $\lambda'^{q}$ and $\left(\underline{P}^{s}_{m,j}\right)^{q}$ to determine the share of the market participants, i.e., $\left(P^{e}_{n,i}\right)^{q}$and $\left(\bar{P}'^{s}_{m,j}\right)^{q}$.

$$\lambda'^{q} = \lambda'^{q-1+\zeta} \tag{5.38}$$

$$\left(\underline{P}^{s}_{m,j}\right)^{q} = \left(P^{s}_{m,j}\right)^{q-1+\zeta} \quad \forall j, \forall m, \forall s \tag{5.39}$$

In step (f), the decisions made by GENCOs and cloud providers are updated similarly based on the sensitivity functions of their associated payoffs in the wholesale market as shown in (5.40) and (5.41) respectively.

$$k^{q+1}_{n,i} = k^{q}_{n,i} + \sigma \cdot \zeta \cdot \frac{\partial R^{q}_{i}}{\partial k^{q}_{n,i}} \quad \forall i, \forall n \tag{5.40}$$

$$k''^{q+1}_{m,j} = k''^{q}_{m,j} + \sigma \cdot \zeta \cdot \frac{\partial R^{q}_{j}}{\partial k''^{q}_{m,j}} \quad \forall i, \forall m \tag{5.41}$$

In step (g), check the observability indicator. If its value is zero its status would be toggled to $\zeta = 1$, and the process will continue to step (c) otherwise go to step (h). The observability indicator shows that the bidding in the wholesale market captures the updated electricity demand in as well as the MCP in the CCM.

In step (h), the convergence conditions (5.42) -(5.44) for all of the participants in both wholesale and cloud computing markets are checked and the equilibrium is established if (5.42)-(5.44) are satisfied. Otherwise, the iteration index and observability indicator ($\zeta$) would be set as $q = q + 1$ and zero respectively and the algorithm continues in next iteration by starting at step (c).

$$\left|\mathbf{k}_{(.)}^{q} - \mathbf{k}_{(.)}^{q-1}\right| \leq \boldsymbol{\epsilon} \tag{5.42}$$

$$\left|\mathbf{k}'^{q}_{(.)} - \mathbf{k}'^{q-1}_{(.)}\right| \leq \boldsymbol{\epsilon} \tag{5.43}$$

$$\left|\mathbf{k}''^{q}_{(.)} - \mathbf{k}''^{q-1}_{(.)}\right| \leq \boldsymbol{\epsilon} \tag{5.44}$$

In the presented framework, each market participant solves a MPEC which is an optimization problem with nonconvex feasibility set. In [5.34], the presence of a unique equilibrium is proved considering certain assumptions. By relaxing some constraints, a unique equilibrium could be achieved as shown in [5.35]. As the MPEC solved by each market participant is a non-convex optimization problem, solving multiple MPEC problems to reach a solution for an EPEC problem is a challenging task. Generally, there might be either one, multiple or no solution for the presented EPEC problems (i.e., Nash equilibrium). In this chapter, the equilibrium point captures the solution for two linked EPEC problems, one in the CCM, and the other in the wholesale market. Therefore, in general, guaranteeing a solution for such problems is not possible due to the non-convex feasibility region of the MPECs formed for each market participant. However, the market regulations could impose additional bounds on the feasibility region of the presented problems that facilitate the convergence and feasibility of the equilibrium. Here, setting the upper and lower bounds for the bidding strategy of the market participants in constraints (5.2), (5.12) and (5.20), would improve the convergence to an equilibrium. At the equilibrium point the market participants are either unwilling to update their bidding strategy or their bidding strategy is bounded. Increasing the number of market participants will further increase the complexity of the proposed problems and increase the solution time. The MCP and LMP are dependent on the bidding strategy of the market participants and as the number of players increases, it is expected that the MCP in the CCM and the LMP in the wholesale market decrease. Furthermore, the payoff of the GENCOs is expected to decrease as the number of GENCOs increase in the wholesale market and their share of the market is decreased.

## 5.5 Case Study

In order to evaluate the efficiency of the proposed approach, two cases were presented. In the first case, three cloud providers were considered in the cloud computing market while in the second case, 10 cloud providers were considered in the wholesale and cloud computing markets.

### 5.5.1 Small Number of Cloud Providers

In order to evaluate the efficiency of the proposed approach, two GENCOs each owning one generator and three cloud providers each owning one data center, are considered in Figure 5-4.

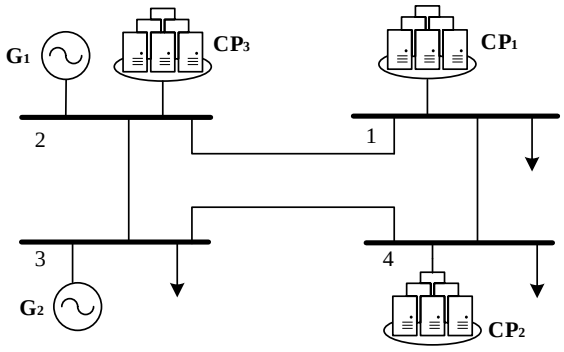


Figure 5-4 Electricity network topology

Table 5-1 Thermal units

| Unit | $a_i$ ($/MW$^2$) | $b_i$ ($/MW) | $c_i$ ($) | $\overline{P_i}$ (MW) |
|---|---|---|---|---|
| $G_1$ | 0.0344 | 8.85 | 0 | 240 |
| $G_2$ | 0.0469 | 8.45 | 0 | 240 |

Table 5-2 Transmission lines characteristics

| Line | From bus | To bus | Impedance (p.u) | Max Flow (MW) |
|---|---|---|---|---|
| 1 | 1 | 2 | 0.0098 | 160 |
| 2 | 2 | 3 | 0.0105 | 250 |
| 3 | 3 | 4 | 0.02 | 190 |
| 4 | 1 | 4 | 0.01 | 130 |

Table 5-3 Power consumption characteristics of cloud providers

| Cloud Provider | $\beta'_j$ (MW/THz$^2$) | $\gamma'_j$ (MW/THz) | $\eta'_j$ (MW) | $\overline{f}_j$ (THz) |
|---|---|---|---|---|
| $CP_1$ | 0.0002 | 0.0668 | 0 | 300 |
| $CP_2$ | 0.0001 | 0.0861 | 0 | 340 |
| $CP_3$ | 0.0003 | 0.0921 | 0 | 150 |

The characteristics of the generation units and transmission lines are shown in Tables 5-1 and 5-2 respectively. The electrical loads on buses 1, 3 and 4 are 70, 90 and 100 MW respectively. The cloud providers' own data centers connected to buses 1, 2, and 4. The capacity of cloud providers $CP_1$, $CP_2$ and $CP_3$ are 34.3 MW, 41.72 MW and 21 MW respectively.

Consequently, $CP_1$ holds 32.8% of the total electricity demand at bus 1, $CP_2$ holds 29.4% of the total electricity demand at bus 4, and $CP_3$ is the only electricity demand on bus 2. The energy consumption of the cloud providers is determined using the quadratic function of the offered frequency in the CCM. The coefficients of the quadratic energy-frequency function are shown in Table 5-3.

The requested frequency from the end users is 350 THz. If a local equilibrium exists, converging to the equilibrium is dependent on the initial bidding strategy. This is mainly because the bi-level optimization problem solved for each market participant – which is further transformed to a MPEC – is a nonlinear and non-convex optimization problem and the solution is dependent on the initial point. In the presented case study, the initial bidding strategy for the market participants in the CCM and the wholesale market is equal to 1 as it is assumed that all market participants bid their marginal costs (i.e., $k = 1, k' = 1, k" = 1$). The lower and upper limits for the bidding vector are 0.5 and 3.0, respectively. The LMP in the wholesale market is initiated as \$20/MWh and the MCP in the CCM is initiated as \$0/THzh. The multiplier $\sigma$ and the threshold for convergence $\varepsilon \in \boldsymbol{\varepsilon}$ are set to $10^{-5}$and $10^{-4}$, respectively. The proposed formulation is solved using MATLAB R2014b on an Intel® Core i5 @3.2 GHz with 8GB of memory using "linprog" function.

The following cases are presented:

Case 1 – Market participants bid at their marginal cost

Case 2 – Market participants bid strategically with no congestion in the electricity network

Case 3 – Market participants bid strategically with congestion in the electricity network

Case 4 – Market participants bid strategically in each market with no coordination among the markets.

*1) Case 1- Market Participants Bid at Their Marginal Cost:*

In this case, all participants bid their marginal costs, therefore, the participants' bidding for offering/consuming the commodities (energy or cloud services) is equal to their marginal cost/benefit and the respective bidding strategies (i.e., $k$, $k'$ , and $k"$ are not updated and fixed to 1). The bidding vectors in the CCM and the wholesale market are calculated as $\mu^{e}_{n,i} = \alpha^{e}_{n,i}$, $\mu'^{s}_{m,j} = \rho_b \cdot \alpha'^{s}_{m,j}$, and $\mu"^{s}_{m,j} = \lambda'$ where $\mu^{e}_{n,i} \in \boldsymbol{\mu}_i$, $\mu"^{s}_{m,j} \in \boldsymbol{\mu}"_j$ and $\mu^{e}_{n,i}$, $\mu'^{s}_{m,j}$, $\mu"^{s}_{m,j}$ are the bids of the segments $e$ and $s$ for GENCOs and cloud providers. Table 5-4 shows the dispatch of the GENCOs and LMPs at each bus along with the expected payoff for each GENCO. Table 5-5 presents the share of each cloud provider in providing the requested frequency along with the expected payoff of each cloud provider and the MCP in the CCM.

As seen in Table 5-4 the awarded dispatch of $G_1$ is more than $G_2$. Furthermore, the marginal cost of $G_1$ is lower than $G_2$, so the payoff of $G_1$ is more than $G_2$. The share of the cloud providers and the MCP in the CCM are shown in Table V. As shown in this table, the awarded dispatch for the cloud provider $CP_1$ is more than that for the cloud providers $CP_2$ and $CP_3$, therefore the payoff for $CP_1$ is higher. In this case, the LMP and MCP are \$20/MWh and \$2.3/THzh respectively. $CP_1$ pays \$392 (19.6 MWh × \$20/MWh = \$392) for consuming 19.6 MWh electricity and earns \$460 (200 THzh × \$2.3/THzh = \$460) for providing 200 THzh in the CCM, which results in \$68 payoff. Similarly, $CP_2$ and $CP_3$ pay \$196 and \$112 in the wholesale market for 9.8 MWh and 5.6 MWh of energy consumed and get paid \$230 and \$115 for processing 100 THzh and 50 THzh in the CCM, respectively. The payoffs for $G_1$ and $G_2$ are \$880 and \$640 for producing 160 MWh and 135 MWh electricity respectively. The solution time for this case is 2.17 seconds.

Table 5-4 Share of GENCOs and LMPs in the wholesale market (Case 1)

| Dispatch of GENCO(MWh) | | LMP at bus (\$/MWh) | | | | Payoff of GENCO (\$) | |
|---|---|---|---|---|---|---|---|
| $G_1$ | $G_2$ | 1 | 2 | 3 | 4 | $G_1$ | $G_2$ |
| 160 | 135 | 20 | 20 | 20 | 20 | 880 | 640 |

Table 5-5 Share of cloud providers and MCP in the CCM (Case 1)

| CPs' share (THzh) | | | MCP (\$/THzh) | Payoff (\$) of CPs | | |
|---|---|---|---|---|---|---|
| $CP_1$ | $CP_2$ | $CP_3$ | 2.3 | $CP_1$ | $CP_2$ | $CP_3$ |
| 200 | 100 | 50 | | 68 | 34 | 3 |

*2) Case2- Market participants bid strategically with no congestion in the electricity network*

In this case, the market participants bid strategically in both markets. The bi-level optimization problem is formed for each market participant and sensitivity functions that capture the participants' payoff functions' sensitivity with respect to the chosen bidding strategy are used to solve the formulated problems. The bidding strategy of the market participants would impact the offered bidding vectors $\mu^{e}_{n,i}, \mu'^{s}_{m,j}$ and $\mu''^{s}_{m,j}$ as shown in (5.3), (5.13), and (5.21). Figure 5-5, Figure 5-6, and Figure 5-7 show the bidding strategies of the cloud providers in the CCM and the wholesale market. As shown in Figure 5-5 the bidding strategies of $CP_1$, $CP_2$ and $CP_3$ are converged to 1.32, 2, and 1.81 respectively. Figure 5-6 represents the bidding strategy of the cloud providers in the wholesale market. In this case, the bidding strategies for $CP_1$, $CP_2$, and $CP_3$ are converged to 2.92, 0.88, and 1 respectively. Figure 5-7 shows that the bidding strategies of $G_1$ and $G_2$ are converged to 2.73 and 3 respectively. As seen in table 5-6, compared with Case 1, the corresponding payoffs increase significantly for $G_1$ and $G_2$, despite the fact that GENCOs produce the same amount of energy. Because of the competition among GENCOs, the LMPs grow from \$20/MWh in Case 1, to \$60/MWh in this case. The share of the cloud providers in providing the requested frequency as well as their bidding strategies in the CCM and wholesale market are shown in Table 5-7. As shown in this table, the cloud providers $CP_1$, $CP_2$ and $CP_3$ bid 1.32, 2 and 1.81 times higher than their respective marginal costs in the CCM. The solution time for this case is 182.084 seconds and the number of iterations to converge is 1557. Compared to Case 1, the bidding of $CP_1$ in the wholesale market is increased to 2.92, while the bidding of $CP_2$ is decreased to 0.88 and $CP_3$ bids its marginal cost. Therefore, because of the strategic bidding of the GENCOs, the cloud providers pay more for the energy purchased from the wholesale market. The payoffs of $G_1$, $G_2$, $CP_1$, $CP_2$ and $CP_3$ are \$7,280, \$6,054, \$1,736, \$868 and \$392 respectively. The energy consumptions of the cloud providers $CP_1$, $CP_2$ and $CP_3$ are 19.6 MWh, 9.8 MWh, and 5.6 MWh respectively, and the energy costs are \$1,176, \$588 and \$336 respectively.

Thus, the revenues of the cloud providers $CP_1$, $CP_2$, and $CP_3$ in the CCM are \$2,912, \$1,456 and \$728 for selling 200 THzh, 100 THzh, and 50 THzh of services respectively. The MCP in the CCM is \$14.56/THzh. The MCP of cloud services in Case 1 is \$2.3/THzh which is further increased to \$14.56/THzh in this case. Therefore, the end-users pay more compared to Case 1, for the same volume of the cloud services.

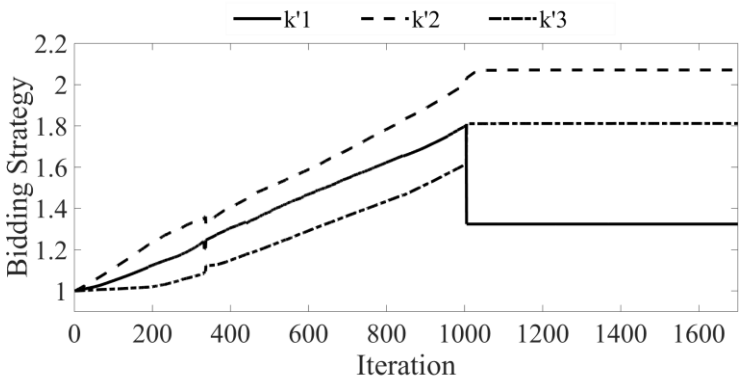


Figure 5-5 Bidding strategy of cloud providers in the CCM

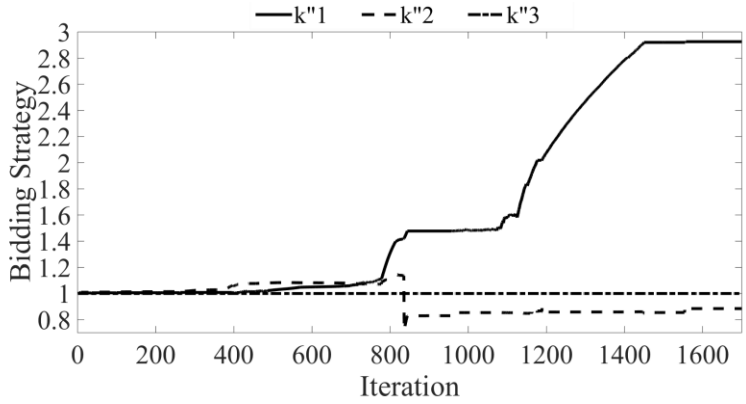


Figure 5-6 Bidding strategy of cloud providers in the wholesale market

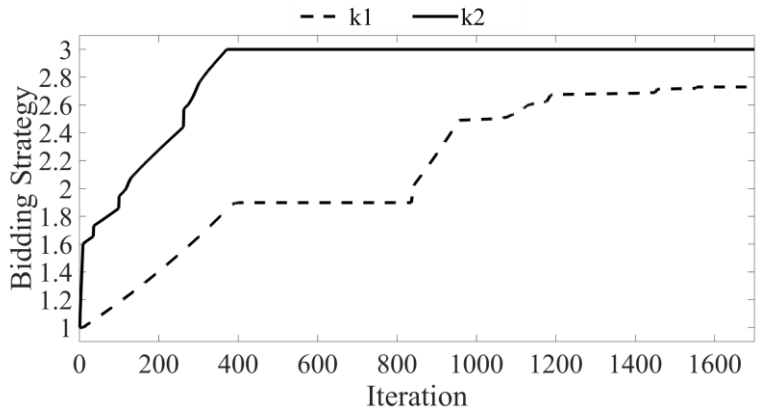


Figure 5-7 Bidding strategy of GENCOs in the wholesale market

Table 5-6 Dispatch of GENCOs and LMPS in the wholesale market (Case 2)

| Dispatch of GENCO (MWh) | | LMP at Bus ($/MWh) | | | | Bidding Strategy of GENCOs | | Payoff of GENCO ($) | |
|---|---|---|---|---|---|---|---|---|---|
| $G_1$ | $G_2$ | 1 | 2 | 3 | 4 | $k_1$ | $k_2$ | $G_1$ | $G_2$ |
| 160 | 135.3 | 60 | 60 | 60 | 60 | 2.7 | 3 | 7280 | 6054 |

Table 5-7 Supplied services by cloud providers and MCP in the CCM (Case 2)

| Cloud Provider's Frequency (THzh) | | | MCP ($/THzh) | Bidding Strategy of Cloud Providers | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | CCM | | | wholesale market | | |
| $CP_1$ | $CP_2$ | $CP_3$ | 14.56 | $k'_1$ | $k'_2$ | $k'_3$ | $k''_1$ | $k''_2$ | $k''_3$ |
| 200 | 100 | 50 | | 1.32 | 2 | 1.81 | 2.92 | 0.88 | 1 |

*3) Case3- Market participants bid strategically with congestion in the electricity network*

In this case, all the participants in CCM and wholesale market bid strategically and line L1 is congested as its capacity is limited to 112 MW. In this case, $CP_1$ and $CP_2$ bid 3 times of their marginal cost while $CP_3$ bid 1.22 of its marginal cost in CCM. Compared to Case 2, the share of G1 in the electricity market is decrease by 65% and the bidding strategy of G1 is increased by 11% as a result of the congestion in line L1. The results of this case are shown in Tables 5-8 and 5-9. As shown in Table 5-8, the dispatches of $G_1$ and $G_2$ are 56 MW and 240 MW respectively. The dispatch of $G_1$ is decreased while the dispatch of $G_2$ is increased compared to those in Case 2. Consequently, the payoff of $G_1$ is decreased to $1,232, while the payoff of $G_2$ is increased to $14,708. Compared to Case 2, the congestion in Line 1 increased the LMPs on buses 1 and 4 where the cloud providers $CP_1$ and $CP_2$ are located and decreased the LMP on bus 2 where the cloud provider $CP_3$ is located.

Table 5-8 Share of GENCOs and LMPs in the wholesale market (Case 3)

| Dispatch of GENCO (MW) | | LMP at Bus ($/MWh) | | | | Bidding Strategy of GENCOs | | Payoff of GENCO ($) | |
|---|---|---|---|---|---|---|---|---|---|
| $G_1$ | $G_2$ | 1 | 2 | 3 | 4 | $k_1$ | $k_2$ | $G_1$ | $G_2$ |
| 56 | 240 | 202.3 | 33 | 81 | 162 | 3 | 3 | 1232 | 14708 |

Table 5-9 Share of cloud providers and MCP in the CCM (Case 3)

| CPs' share (THzh) | | | MCP ($/THzh) | Bidding Strategy of CPs | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | CCM | | | wholesale market | | |
| $CP_1$ | $CP_2$ | $CP_3$ | 56.8 | $k'_1$ | $k'_2$ | $k'_3$ | $k''_1$ | $k''_2$ | $k''_3$ |
| 100 | 100 | 150 | | 3 | 3 | 1.22 | 1.45 | 1.61 | 1 |

The payoffs for cloud provider $CP_1$, $CP_2$, and $CP_3$ are $3,611 and $3,724, and $7,272 respectively. Comparing with Case 2, $G_2$ and the cloud providers benefit from congestion while the payoff of $G_1$ is decreased. Total

payoff of cloud providers, in this case, is \$14,607 which is increased compared to Case 2. Therefore, in this case, the end users should pay more for the cloud services as the MCP in CCM is increased to \$53.1/THzh. The number of iterations to converge in this case is 1703 and the solution time is 218.807 seconds.

4) *Case4- Market participants bid strategically in each market with no coordination among the markets.*

In this case, the interactions among cloud computing and wholesale markets are ignored to highlight the merit of capturing the interdependence among these markets. Here, the cloud providers in the wholesale market cannot change their strategy iteratively and they are bounded by the consequences of their adopted bidding strategies in the CCM. The cloud providers compete in the CCM to offer cloud services to the end users considering an initial price of the electricity (\$20/MWh) and the outcome of this competition (i.e., electricity demand) sets boundary conditions for the bidding of the demand in the wholesale market. Table 5-10 represents the results of the competition in the CCM. Considering the realized electricity demand in the CCM, the cloud providers bid in the wholesale market and the outcomes of the competition in the wholesale market are shown in Table 5-11.

Table 5-10 Share of the cloud providers and MCP in the CCM (Case 4)

| CPs' share (THzh) | | | MCP (\$/THzh) | Bidding Strategy of CPs | | |
|---|---|---|---|---|---|---|
| | | | | CCM | | |
| $CP_1$ | $CP_2$ | $CP_3$ | 6.1 | $k'_1$ | $k'_2$ | $k'_3$ |
| 200 | 100 | 50 | | 2.67 | 3 | 2.3 |

Table 5-11 Share of the GENCOs and LMPs in the wholesale market (Case 4)

| Dispatch of GENCO (MW) | | LMP at Bus (\$/MWh) | | | | Bidding Strategy of GENCOs | | Payoff of GENCO (\$) | |
|---|---|---|---|---|---|---|---|---|---|
| $G_1$ | $G_2$ | 1 | 2 | 3 | 4 | $k_1$ | $k_2$ | $G_1$ | $G_2$ |
| 160 | 135 | 58 | 58 | 58 | 58 | 2.88 | 3 | 6960 | 5770 |

Considering \$20/MWh as the initial price for the electricity, $CP_1$, $CP_2$, $CP_3$ are expected to pay \$392, \$196, and \$112 for consuming 19.6 MWh, 9.8 MWh, and 5.6 MWh energy in the wholesale market respectively. However, after the LMPs are procured in the wholesale market, $CP_1$, $CP_2$, and $CP_3$ should pay \$1,136.8, \$568.4, and \$324.8 respectively. Considering \$20/MWh as the price for the purchased energy, the cloud providers $CP_1$, $CP_2$, and $CP_3$ expect their payoffs to be \$828, \$414, and \$193 respectively; however, after

participating in the wholesale market using the offered frequency services in the CCM, their actual payoffs are reduced to \$83.2, \$41.6, \$-19.8 respectively. The solution time for this case is 188.95 seconds, the number of iterations to converge in the CCM is 451, and the number of iterations to converge in the wholesale market is 1219.

### 5.5.2 Large Number of Cloud providers

In order to investigate the impact of larger number of players in the competition, 10 cloud providers were considered in the electricity network shown in Figure 5-4. The coefficients of the quadratic energy-frequency functions of the cloud providers, along with their locations in the power system, are shown in Table 5-12. Here, four cases, Case 1-4 were considered. The requested frequency of the end users is 350 THz. The market shares of the cloud providers in the CCM, their bidding strategies in the CCM and the wholesale market, their payoffs, and the MCP in the CCM are shown in Table 5-13 for Cases 1-4. The dispatch of the GENCOs, their bidding strategies in the wholesale market, their payoffs, and the LMP in the wholesale market are shown in Table 5-14.

In Case 1, the cloud providers $CP_1$, $CP_3$, $CP_5$, $CP_7$, $CP_8$, and $CP_{10}$ offer cloud services to the end users, as their bid their marginal costs. The dispatches of GENCOs $G_1$ and $G_2$ are 160, and 128 MW respectively. By bidding strategically in Case 2 the cloud providers $CP_1$, $CP_3$, $CP_5$, $CP_6$, $CP_7$, and $CP_{10}$ benefit from offering cloud services. Compared to Case 1, it is observed that the payoffs of the cloud providers were increased considerably. For instance, the payoff of cloud provider $CP_1$ increases from \$44 to \$876, for providing the same amount of cloud services. Despite the same dispatch of the GENCOs, the corresponding payoffs were increased. The MCP and LMP grow from \$2.12/THzh, and \$20/MWh in Case 1 to \$13.8/THzh and \$60/MWh in Case 2, respectively. In Case 3, where line L1 is congested, it is seen that, the cloud providers $CP_1$ and $CP_5$ have limited opportunities to offer cloud services in the CCM. In this case, the shares of the cloud providers $CP_6$ and $CP_7$ increase from 24, and 62 THzh to 48, and 180 THzh respectively, and the cloud provider $CP_8$ offers 50 THzh in the CCM. Ignoring the interaction between the electricity and cloud computing markets in

Case 4 would result in considerable decrease in the payoffs of the market participants and economical loss for the cloud providers $CP_3$, $CP_6$, and $CP_8$.

Table 5-12 Characteristics of cloud providers

| Cloud Provider | Location | $\beta'_j$ (MW/THz$^2$) | $\gamma'_j$ (MW/THz) | $\eta'_j$ (MW) | $\overline{f}_j$ (THz) |
|---|---|---|---|---|---|
| $CP_1$ | Bus 1 | 0.0002 | 0.0668 | 0 | 300 |
| $CP_2$ | Bus 4 | 0.0001 | 0.0861 | 0 | 340 |
| $CP_3$ | Bus 2 | 0.0003 | 0.0921 | 0 | 150 |
| $CP_4$ | Bus 1 | 0.0002 | 0.1124 | 0 | 90 |
| $CP_5$ | Bus 1 | 0.0018 | 0.0245 | 0 | 63 |
| $CP_6$ | Bus 2 | 0.0002 | 0.0359 | 0 | 72 |
| $CP_7$ | Bus 3 | 0.0004 | 0.0182 | 0 | 270 |
| $CP_8$ | Bus 3 | 0.0007 | 0.0458 | 0 | 135 |
| $CP_9$ | Bus 3 | 0.0007 | 0.0227 | 0 | 153 |
| $CP_{10}$ | Bus 4 | 0.0009 | 0.0497 | 0 | 108 |

Table 5-13 Outcomes for the cloud providers

| | Market share in the CCM(THzh) | | | | Payoff ($) | | | | Bidding strategy in the CCM | | | | Bidding strategy in the wholesale market | | | | MCP in the CCM ($/THzh) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Case | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| $CP_1$ | 100 | 100 | 0 | 100 | 44 | 876 | 0 | 56 | 1 | 1 | 1.017 | 1 | 1 | 1.126 | 3 | - | | | | |
| $CP_2$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1.21 | 0.5 | - | | | | |
| $CP_3$ | 16.5 | 50 | 0 | 50 | 2.7 | 354 | 0 | -47.4 | 1 | 0.685 | 1 | 1.3 | 1 | 1.05 | 0.5 | - | | | | |
| $CP_4$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1.02 | 0.5 | - | | | | |
| $CP_5$ | 21 | 42 | 0 | 50 | 6.3 | 103.3 | 0 | -65 | 1 | 1 | 1 | 1 | 1 | 1.02 | 0.5 | - | | | | |
| $CP_6$ | 0 | 24 | 48 | 0 | 0 | 220.3 | 1259 | 0 | 1 | 1.182 | 1.034 | 1.08 | 1 | 1.01 | 0.5 | - | 2.12 | 13.8 | 35.3 | 4.35 |
| $CP_7$ | 120.7 | 62 | 180 | 65 | 99.7 | 595.2 | 3904 | 85.8 | 1 | 3 | 1.021 | 1.11 | 1 | 1.03 | 3 | - | | | | |
| $CP_8$ | 45 | 0 | 50 | 45 | 7.2 | 0 | 1362 | -8.6 | 1 | 1 | 3 | 1 | 1 | 0.52 | 0.631 | - | | | | |
| $CP_9$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1.024 | 1 | 1 | 1.07 | 0.644 | - | | | | |
| $CP_{10}$ | 46.8 | 72 | 72 | 40 | 8.5 | 25.9 | 502 | 10.46 | 1 | 0.513 | 1.157 | 1.05 | 1 | 1.21 | 0.622 | - | | | | |

Table 5-14 Outcomes for GENCOs in wholesale market

| | Dispatch of GENCO (MW) | | | | Payoff of GENCO ($) | | | | Bidding Strategy of GENCOs | | | | LMP on Bus ($/MWh) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | | | 3 | | | | | |
| Case | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | Bus 1 | Bus 2 | Bus 3 | Bus 4 | 4 | |
| $G_1$ | 160 | 160 | 85.2 | 240 | 880 | 7280 | 3631 | 8880 | 1 | 2.8 | 3 | 3 | 20 | 60 | 149.2 | 54 | 81 | 126.5 | 54 | |
| $G_2$ | 128 | 128 | 206 | 48.1 | 640 | 5760 | 12889 | 2020.2 | 1 | 3 | 3 | 3 | | | | | | | | |

## 5.6 Summary

In this chapter, the strategic behavior of the cloud providers in the wholesale market and CCM is investigated. In the wholesale market, the cloud providers bid to minimize their energy cost while GENCOs maximize their profit. The cloud providers maximize their revenue in the CCM by bidding for the cloud

services provided to the end-users. The competition for the cloud providers in the wholesale market and CCM are presented as a dynamic game with complete information. In the proposed structure, the interdependence among the price of electricity and the price of the cloud services is captured. The bidding strategy chosen by the market participants is procured by solving a bi-level optimization problem in which the upper-level problem is a profit maximization problem while the lower-level problem represents the market clearing process. Sensitivity functions with respect to the adopted bidding strategies for the market participants' payoff functions were developed to solve the respective bi-level optimization problems. The presented case study verifies the efficiency of proposed structure. It is shown that once the market participants in CCM and the wholesale market bid their marginal costs, their payoffs are lower than the case in which they strategically bid for the offered and purchased services. It is shown that ignoring the interdependence between the CCM and the wholesale market will lead to a considerable loss in payoff for the cloud providers. Furthermore, it is shown that because of congestion in the electricity network, the cloud providers whose energy consumption would reduce the congestion will pay less for their energy in the wholesale market and therefore, they have higher payoffs compared to the other cloud providers. Furthermore, the congestion in the electricity network would limit the less expensive generation units to serve the demand. Therefore, units that are more expensive provide more energy in the wholesale market and their payoff will increase accordingly. The proposed solution methodology can be further extended to address the non-cooperative dynamic game with incomplete information by considering different "types" for the market participants in the CCM and the wholesale market.

## 5.7 References


[5.1] Y. Chen et al., "Integrated management of application performance, power and cooling in data centers," in *Proc. IEEE Netw. Oper. Manag. Symp. (NOMS),* Osaka, Japan, 2010, pp. 615–622.

[5.2] J. Heo et al., "OptiTuner: On performance composition and server farm energy minimization application," *IEEE Trans. Parallel Distrib. Syst.,* vol. 22, no. 11, pp. 1871–1878, Nov. 2011.

[5.3] M. Lin, Z. Liu, A. Wierman, and L. L. H. Andrew, “Online algorithms for geographical load balancing,” in *Proc. Int. Green Comput. Conf. (IGCC),* San Jose, CA, USA, 2012, pp. 1–10

[5.4] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. H. Andrew, “Geographical load balancing with renewables,” *ACM SIGMETRICS Perform. Eval. Rev.,* vol. 39, no. 3, pp. 62–66, 2011.

[5.5] W. Baek and T. M. Chilimbi, “Green?: A framework for supporting energy-conscious programming using controlled approximation,” in *Proc. ACM SIGPLAN Conf. Program. Lang. Des. Implement.* Toronto, ON, Canada, 2010, pp. 198–209.

[5.6] K. Wang, M. Lin, F. Ciucu, A. Wierman, and C. Lin, “Characterizing the impact of the workload on the value of dynamic resizing in data centers,” *Perform. Eval.,* vols. 85–86, pp. 1–18, Jul. 2015.

[5.7] J. Hamilton, “Cooperative expendable micro-slice servers (CEMS): Low cost, low power servers for Internet-scale services,” in *Proc. 4th Biennial Conf. Innov. Data Syst. Res. (CIDR),* Asilomar, CA, USA, Jan. 2009, pp. 770–791.

[5.8] R. Bianchini and R. Rajamony, “Power and energy management for server systems,” *Computer,* vol. 37, no. 11, pp. 68–76, Nov. 2004.

[5.9] D. Brooks and M. Martonosi, “Dynamic thermal management for highperformance microprocessors,” in *Proc. HPCA 7th Int. Symp. High Perform. Comput. Archit.,* Monterrey, Mexico, 2001, pp. 171–182.

[5.10] J. Li, Z. Bao, and Z. Li, “Modeling demand response capability by Internet data centers processing batch computing jobs,” *IEEE Trans. Smart Grid,* vol. 6, no. 2, pp. 737–747, Mar. 2015.

[5.11] Y. Yao, L. Huang, A. Sharma, L. Golubchik, and M. Neely, “Data centers power reduction: A two time scale approach for delay tolerant workloads,” in *Proc. IEEE INFOCOM*, Orlando, FL, USA, 2012, pp. 1431–1439.

[5.12] M. Ghamkhari and H. Mohsenian-Rad, “Energy and performance management of green data centers: A profit maximization approach,” *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 1017–1025, Jun. 2013

[5.13] *Amazon EC2 Pricing*. Accessed: Mar. 2017. [Online]. Available: http://aws.amazon.com/ec2/ pricing/

[5.14] *Amazon Simple Queue Service (Amazon SQS).* Accessed: Mar. 2017. [Online]. Available: http://aws.amazon.com/sqs/

[5.15] *Google App Engine Pricing.* Accessed: Mar. 2017. [Online]. Available: http://cloud.google.com/pricing/

[5.16] Y. Feng, B. Li, and B. Li, “Price competition in an oligopoly market with multiple IaaS cloud providers,” *IEEE Trans. Comput*., vol. 63, no. 1, pp. 59–73, Jan. 2014.

[5.17] J. Anselmi, U. Ayesta, and A. Wierman, “Competition yields efficiency in load balancing games,” *Perform. Eval*. vol. 68, no. 11, pp. 986–1001, 2011.

[5.18] D. Ardagna, B. Panicucci, and M. Passacantando, “Generalized Nash equilibria for the service provisioning problem in cloud systems,” *IEEE Trans. Services Comput*., vol. 6, no. 4, pp. 429–442, Oct./Dec. 2012.

[5.19] D. Acemoglu and A. Ozdaglar, “Competition and efficiency in congested markets,” *Math. Oper. Res*., vol. 32, no. 1, pp. 1–31, 2007.

[5.20] Y. Fu, J. Chase, B. Chun, S. Schwab, and A. Vahdat, “SHARP: An architecture for secure resource peering,” *ACM SIGOPS Oper. Syst. Rev.,* vol. 37, no. 5, pp. 133–148, 2003.

[5.21] K. Lai, L. Rasmusson, E. Adar, L. Zhang, and B. A. Huberman, “Tycoon: An implementation of a distributed, market-based resource allocation system,” *Multiagent Grid Syst.,* vol. 1, no. 3, pp. 169–182, 2005

[5.22] A. AuYoung, B. Chun, A. Snoeren, and A. Vahdat, “Resource allocation in federated distributed computing infrastructures,” in *Proc. 1st Workshop Oper. Syst. Archit. Support Ondemand IT Infrastruct. (OASIS),* Boston, MA, USA, Oct. 2004, pp. 1–10.

[5.23] D. E. Irwin *et al*., "Sharing networked resources with brokered leases," in *Proc. USENIX Annu. Tech. Conf. (USENIX)*, Boston, MA, USA, Jun. 2006, p. 18.

[5.24] *PLANETLAB, an Open Platform for Developing, Deploying, and Accessing Planetary-Scale Services.* Accessed: Mar. 2017. [Online]. Available: https://www.planet-lab.org/

[5.25] R. Buyya, C. S. Yeo, and S. Venugopal, "Market-oriented cloud computing: Vision, hype, and reality for delivering IT services as computing utilities," *in Proc. 10th IEEE Int. Conf. High Perform. Comput. Commun*. Dalian, China, 2008, pp. 5–13.

[5.26] M. Shahidehpour, H. Yamin, and Z. Li, *Market Operations in Electric Power Systems*. New York, NY, USA: Wiley, 2002.

[5.27] R. W. Ferrero, J. F. Rivera, and S. M. Shahidehpour, "Application of games with incomplete information for pricing electricity in deregulated power pools," *IEEE Trans. Power Syst.,* vol. 13, no. 1, pp. 184–189, Feb. 1998

[5.28] S. A. Gabriel, J. Zhuang, and R. Egging, "Solving stochastic complementarity problems in energy market modeling using scenario reduction," *Eur. J. Oper. Res.,* vol. 197, no. 3, pp. 1028–1040, 2009.

[5.29] T. Li and M. Shahidehpour, "Strategic bidding of transmissionconstrained GENCOs with incomplete information," *IEEE Trans. Power Syst.,* vol. 20, no. 1, pp. 437–447, Feb. 2005.

[5.30] S. J. Wang, S. M. Shahidehpour, D. S. Kirschen, S. Mokhtari, and G. D. Irisarri, "Short-term generation scheduling with transmission and environmental constraints using an augmented Lagrangian relaxation," *IEEE Trans. Power Syst*., vol. 10, no. 3, pp. 1294–1301, Aug. 1995

[5.31] A. Qureshi, R. Weber, H. Balakrishnan, J. Guttag, and B. Maggs, "Cutting the electric bill for Internet-scale systems," *ACM SIGCOMM Comput. Commun. Rev*., vol. 39, no. 4, pp. 123–134, 2009.

[5.32] *Power-Consumption Scaling With Clockspeed and Vcc for the i7-2600K.* Accessed: Mar. 2017. [Online]. Available: http://forums.anandtech.com/ showthread.php?t=2195927

[5.33] S. D. Manshadi and M. E. Khodayar, “A hierarchical electricity market structure for the smart grid paradigm,” *IEEE Trans. Smart Grid*, vol. 7, no. 4, pp. 1866–1875, Jul. 2016.

[5.34] H. D. Sherali, “A multiple leader Stackelberg model and analysis,” *Oper. Res.*, vol. 32, no. 2, pp. 390–404, Apr. 1984.

[5.35] V. DeMiguel and H. Xu, “A stochastic multiple-leader Stackelberg model: Analysis, computation, and application,” *Oper. Res.*, vol. 57, no. 5, pp. 1220–1235, Sep. 2009.

[3.36] Razzaghi, Pouria, Ehab Al Khatib, and Yildirim Hurmuzlu. "Nonlinear dynamics and control of an inertially actuated jumper robot." Nonlinear Dynamics, pp. 1-16, 2019).

Chapter 6

# OPERATION OF DISTRIBUTION NETWORKS WITH VOLATILE SUPPLY AND CONTROLLABLE DATA CENTER DEMAND

The proposed framework is used by the DSO to determine the demand of distribution network in the day ahead market based on the forecasted day-ahead hourly and intra-hourly electricity prices. The data centers process batch computing workloads that arrive at certain periods. DSO operate the generation and controllable demand assets including the data centers to ensure the economic and security of the distribution network while satisfying the workload processing requirements for data centers. As shown in Figure 6-1, DSO communicates with independent system operator (ISO) to determine the required power supplied by the bulk power system.

The uncertainties in the generation and demand introduces economic risk in the distribution network operation and was captured by introducing effective scenarios. Capturing the risk in the distribution network operation framework enables the DSO to avoid over-conservative operation strategies. The proposed framework incorporates risk aversion by constraining the volatility of the expected operation cost through conditional value at risk (CVaR) assessment. The rest of this chapter is organized as follows: Section 6.1 describes the problem formulation. Sections 6.2 and 6.3 present the case study and summary respectively.

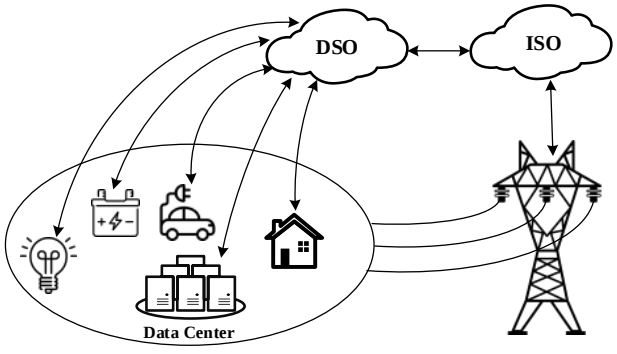


Figure 6-1 Active distribution network with data centers

## 6.1 List of Symbols

**Indices:**

| | |
|---|---|
| $d$ | Index of electrical loads other than data center |
| $h$ | Index of intra-hour |
| $j$ | Index of workload |
| $m$ | Index of module in a data center |
| $n$ | Index of interval |
| $s$ | Index of scenario |
| $t$ | Index of hour |
| $w$ | Index of wind unit |

**Binary variables:**

| | |
|---|---|
| $I_n^j$ | Processing state of the workload |
| $k_n^m$ | Operation status of module $m$ of data center |
| $y_t^m / z_t^m$ | Wake up/Shut down indicator of data center |
| $y'^j_t / z'^{mj}_t$ | Initiation/completion indicator of the workload |

**Real variables:**

| | |
|---|---|
| $C'^s$ | Operation cost of scenario $s$ [$] |
| $C$ | Expected operation cost of the distribution network [$] |
| $CVaR$ | Conditional value at risk [$] |
| $f''^{s,j}_n$ | Frequency provided for processing workload $j$ of scenario $s$ at interval $n$ [GHz] |
| $P'^{s,in}_t$ | Hourly demand of the distribution network at hour $t$ in scenario $s$ [MW] |
| $P_n^{s,g}$ | Intra-hour demand of the distribution network at interval $n$ in scenario $s$ [MW] |
| $P_n^{s,m}$ | Demand of module $m$ of data center at interval $n$ in scenario $s$ [MW] |
| $sd_n^m / su_n^m$ | OFF/ON time of module $m$ [Minutes] |
| $u_n^{s,m}$ | Utilization factor of module $m$ |
| $VaR$ | Value at risk [$] |

| | |
|---|---|
| $\lambda^s$ | Auxiliary variable |

**Parameters:**

| | |
|---|---|
| $f^m$ | Maximum frequency of module $m$ [GHz] |
| $f'^{s,j}_{req}$ | Processing capability required to handle workload $j$ in scenario $s$ [GHzh] |
| $mf^m/mn^m$ | Initially OFF/ON periods of module $m$ [Minutes] |
| $md^m/mu^m$ | Minimum OFF/ON time of data center [Minutes] |
| $P^m_{idle}$ | Power consumption of module $m$ in idle mode [MW] |
| $P^m_{act}$ | Dynamic power consumption of module $m$ with full utilization [MW] |
| $P^{s,d}_n$ | Distribution network demand other than data centers [MW] |
| $P^{s,w}_n$ | Forecasted dispatch of wind generation |
| $N^j_{st}$ | Start time for processing the workload |
| $N^j_{dl}$ | Deadline for processing the workload |
| $NH$ | Number of intervals in an hour |
| $NT$ | Total number of intervals in the operation horizon |
| $prob^s$ | Probability of scenario |
| $\alpha$ | Threshold for CVaR |
| $\omega$ | Trade-off parameter |
| $\rho'^s_t$ | Day ahead electricity price at hour $t$ in scenario $s$ [$/MWh] |
| $\rho^s_n$ | Intra-hour electricity price, at interval $n$ in scenario $s$ [$/MWh] |
| $\psi$ | Set of workloads that require being processed continuously |

## 6.2 Problem Formulation

The objective of DSO is to minimize the day-ahead hourly and intra-hourly demand charges in the electricity network. The objective function is the expected operation cost of the distribution network in the day-ahead market considering the respective uncertainties in the intra-hour local generation resources and non-controllable loads, as well as the uncertainty in the intra-hour price of electricity. The data center is

considered as a controllable load which can be further coordinated with the renewable generation resources and non-controllable demands in the distribution network. The uncertainties in the receiving workloads for the data centers are captured for the load management of data centers. Furthermore, the proposed model incorporates the risk associated with the decisions made by DSO through CVaR assessment [6.7].

Value at risk (VaR), is a common tool for measuring and managing risk, which is measured in price units under certain confidence level $\alpha \in (0,1)$ (usually 95% or 99%).It represents the lowest operation cost, for which the probability of procuring the total operation cost higher than this value is lower than $(1-\alpha)$. While VaR provides overall vision on risk assessment, relying on it can be very misleading, as it does not measure the maximum potential operation cost in the lower $(1-\alpha)$ percentile. Particularly, for the discontinuous distribution of VaR or distributions other than normal distribution, the operation cost expected to be greater than the VaR amount, in $(1-\alpha)$ percentage might lead to extreme operation cost due to the fat tail in distribution profile [6.8].

Unlike, VaR, CVaR measures the conditional expected operation cost of the distribution network above the VaR, in the lower $(1-\alpha)$ percentile of the probability distribution function of the operation cost. Furthermore, CVaR is a convex coherent risk averse function that allows for forming convex stochastic optimization problems [6.9].

In order to determine the day-ahead and real-time load of the distribution network with controllable data center demand, the problem (6.1)-(6.27) is formulated. The objective function is shown in (6.1), where the first term is the weighted expected operation cost of the distribution network and the second term is the weighted CVaR that determines the risk associated with the operation cost of the distribution network. The trade-off parameter $\omega$ takes values between 0 and 1. The expected operation cost of the distribution network is calculated in (6.2), where the operation cost of each scenario is presented in (6.3). The operation cost in each scenario consists of two terms, the first term captures the hourly operation cost in each scenario, and the second term addresses the intra-hourly operation cost. In (6.4), the auxiliary variable $\lambda^s$ is positive if the expected operation cost of any scenario exceeds the VaR. In (6.5), CVaR is calculated considering the distribution network operation cost in all scenarios. In (6.6), the amount of day-ahead hourly demand is equal to the total intra-hourly demand. The balance among the demand and supply is ensured by (6.7). The power consumption of a module of data center is calculated in (6.8), where the first term, represents its power

consumption in idle mode, and the second term captures its dynamic power consumption in active mode, that is dependent on the utilization factor. Constraint (6.9) limits the utilization factor of the data center modules considering their operation state. In (6.10), the total frequency required for processing the workloads at interval *n* determines the utilization factor of the data center modules. In (6.11), the sum of provided frequency to process each workload during the processing period, should be equal to processing capability required for that workload. In (6.12) and (6.13), the workloads are enforced to be processed within their respective deadlines. In (6.14), the frequency provided for processing the workloads at interval *n* takes value if the execution status is 1 and $\epsilon$ is a small scalar. Constraints (6.15) and (6.16) capture the relationship between workload initiation and completion variables and the status of the workload process. Constraint (6.17) enforces the continuous processing of the workload if the workload requires to be processed without any interruption. Constraints (6.18)-(6.21) and (6.22)-(6.25) enforce the minimum OFF/ON time of the data center modules. Constraints (6.26) and (6.27) capture the relationship between the operating state of the data center modules and their respective startup/shutdown indicators.

$$\min(1-\omega)\cdot C+\omega\cdot CVaR \tag{6.1}$$

*s.t.*

$$C=\sum_s prob^s\cdot C'^s \tag{6.2}$$

$$C'^s=\sum_t\left[\rho'^s_t\cdot P'^{s,g}_t+\sum_{n\in N_t}\frac{1}{NH}\cdot\left(\rho^s_n\cdot\left(P^{s,g}_n-P'^{s,g}_t\right)\right)\right]\quad N_t=\{n|n=NH\times(t-1)+h\ and\ h=1,2,\dots,NH\} \tag{6.3}$$

$$C'^s-\lambda^s-VaR\le 0 \tag{6.4}$$

$$CVaR=VaR+\frac{1}{1-\alpha}\sum_s prob^s\cdot\lambda^s \tag{6.5}$$

$$P'^{s,g}_t=\sum_{n\in N_t}\frac{1}{NH}\cdot P^{s,g}_n \tag{6.6}$$

$$P^{s,in}_n+P^{s,w}_n=P^{s,d}_n+\sum_m P^{s,m}_n \tag{6.7}$$

$$P^{s,m}_n=P^m_{idle}\cdot k^m_n+P^m_{act}\cdot u^{s,m}_n \tag{6.8}$$

$$u^{s,m}_n\le k^m_n \tag{6.9}$$

$$\sum_m u^{s,m}_n\cdot f^m=\sum_j f''^{s,j}_n \tag{6.10}$$

$$\sum_n f''^{s,j}_n = NH \cdot f'^{s,j}_{req} \tag{6.11}$$

$$\sum_{n=N^j_{st}}^{N^j_{st}+N^j_{dl}} I^j_n \le N^j_{dl} \tag{6.12}$$

$$\sum_{n=0}^{N^j_{st}} I^j_n + \sum_{n=N^j_{st}+N^j_{dl}}^{NT} I^j_n \le 0 \tag{6.13}$$

$$I^j_n \cdot \epsilon \le f''^{s,j}_n \le I^j_n \cdot \sum_m f^m \tag{6.14}$$

$$y'^j_n - z'^j_n = I^j_n - I^j_{n-1} \tag{6.15}$$

$$y'^j_n - z'^j_n \le 1 \tag{6.16}$$

$$\sum_n y'^j_n = 1, \quad \forall j \in \psi \tag{6.17}$$

$$sd^m_n \le NH \cdot mf^m \cdot (1 - k^m_n) \tag{6.18}$$

$$sd^m_n - sd^m_{n-1} \le 1 \tag{6.19}$$

$$sd^m_n - sd^m_{n-1} \ge 1 - NH \cdot (mf^m + 1) \cdot k^m_n \tag{6.20}$$

$$sd^m_n \ge NH \cdot md^m \cdot y^m_{n+1} \tag{6.21}$$

$$su^m_n \le NH \cdot mn^m \cdot k^m_n \tag{6.22}$$

$$su^m_n - su^m_{n-1} \le 1 \tag{6.23}$$

$$su^m_n - su^m_{n-1} \ge NH \cdot \left((mn^m + 1) \cdot k^m_n - mn^m\right) \tag{6.24}$$

$$su^m_n \ge NH \cdot mu^m \cdot z^m_{n+1} \tag{6.25}$$

$$y^m_n + z^m_n \le 1 \tag{6.26}$$

$$y^m_n - z^m_n = k^m_n - k^m_{n-1} \tag{6.27}$$

## 6.3 Case Study

In this section, to evaluate the efficiency of the proposed framework, a data center with two modules is considered as controllable loads in the distribution network. Table 6-1 represents the characteristics of the data center modules, including the maximum frequency provided by the data center modules, as well as the idle and dynamic power consumptions. Table 6-2 represents the workload characteristics including the required time for processing of the workloads, as well as the receiving time and the deadline for the workloads. The wind unit has 9 MW generation capacity. The mean value of wind speed is assumed to be 10

m/s, and the shape parameter is considered as 2.1. The intra-hourly electricity prices and 15-minute distribution network demand were shown in Figure 6-2 [6.10]. The mean of intra-hourly electricity prices was considered as the day-ahead hourly price of electricity.

Table 6-1 Characteristics of the data center modules

| Module | Frequency (THz) | Idle power (MW) | Max Dynamic power (MW) | Min on time (hour) | Min off time (hour) |
|---|---|---|---|---|---|
| 1 | 900 | 5 | 13 | 2 | 1 |
| 2 | 800 | 3 | 15 | 2 | 1 |

Table 6-2 Workloads characteristics

| Workload | Continuous workload | Arrival time | Process requirement [GHzh] | Deadline (hours) |
|---|---|---|---|---|
| 1 | No | 3:30 | 2700 | 20 |
| 2 | No | 7:15 | 4400 | 15 |
| 3 | Yes | 11:45 | 3000 | 12 |

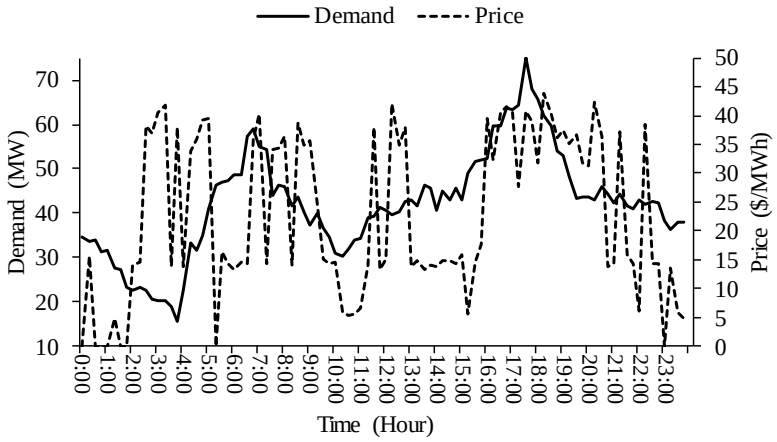


Figure 6-2 Intra-hourly demand and price of electricity

Monte-Carlo Simulation (MCS) is utilized to generate a large number of scenarios to capture the uncertainties in the operation horizon. To incorporate the forecast errors of the electricity demand and the intra-hourly electricity price, the normal distribution function is used with the mean value equal to the forecasted intra-hourly demand and the price of electricity at each interval respectively. The standard deviation is set to 5% of the mean values. The uncertainty in wind speed is addressed using Weibull probability distribution function, and the wind speed is used to determine the wind generation using wind turbine speed-power curve. In this case study, 3000 scenarios were generated, and fast backward/forward scenario reduction technique is used to reduce the number of effective scenarios to 13.

Three cases were considered as follows:

Case 1 – Uncoordinated operation of data centers and distribution network

Case 2 – Coordinated operation of data centers and distribution network - Deterministic solution

Case 3 – Coordinated operation of data centers and distribution network - Stochastic solution

### 6.3.1 Case 1- Uncoordinated Operation of Data Centers and Distribution Network

In this case, it is assumed that the data center operator processes the workloads immediately without any workload management strategy. The objective is to minimize the power consumption associated with processing the workloads. As the data center is not operated in coordination with the distribution network, the DSO cannot leverage the flexibility of data centers to minimize the operation cost of the distribution network. As seen in Table 6-2, workload 1 is requested at 3:30, which starts to be processed immediately until 6:45. Workloads 2 and 3 start to be processed at 7:15, and 11:45 and their processes were finished at 16:00, and 15:30 respectively. In this case, all workloads were assigned to module 1 and the total energy consumption of data centers is 202.14 MWh. Considering the intra-hour price of electricity illustrated in Figure 2, the data center's operation cost is $4,500 and the operation cost of the distribution network is $21,255.

### 6.3.2 Case 2- Coordinated Operation of Data centers and Distribution Network: Deterministic Solution

In this case, DSO leverages the proposed framework to manage the data center demand and to minimize the total operation cost of distribution network while ignoring the uncertainties in the operation horizon. In this case, all workloads have been processed within their corresponded deadlines. Workload 1, has been requested at 3:30, and its execution process is interrupted three times. It is processed at 5:15-6:30, 9:45- 10:00, 11:15-11:30, and 22:30-23:00. Workload 2, which arrives at 7:15, is processed at 9:45-12:15 and 21:30-22:00. Workload 3 that is requested at 11:45, is processed at 20:45-23:45. As workload 3 is required to be processed continuously, it is seen that there is no interruption in its process.

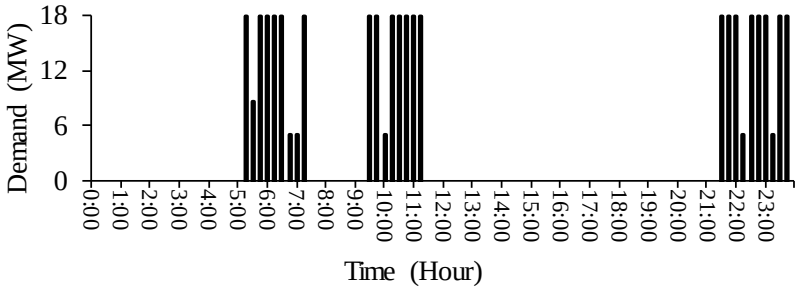


Figure 6-3 Power consumption of module 1

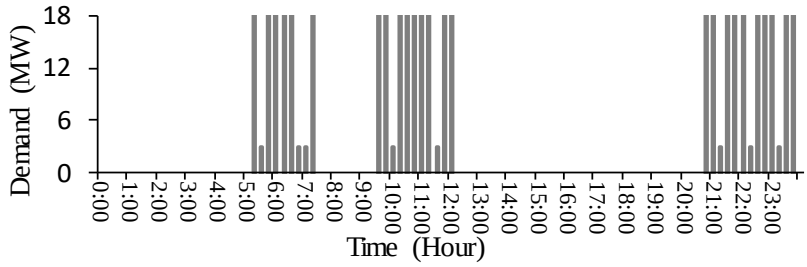


Figure 6-4 Power consumption of module 2

In Figures 6-3 and 6-4, the power consumption profile of the module 1 and module 2 of data center were depicted, respectively. Here, both modules are shut down twice. Module 1 is in idle mode in 6:45-7:15, 10:00-10:15, 22:15-22:30, and 23:15-23:30 respectively. Similarly, module 2 is in idle mode in 5:30-5:45, 7:00-7:15, 10:00-10:15, 11:30-11:45, 21:15-21:30, 22:15-22:30, and 23:15-23:30 respectively. At these periods data center modules, do not process any workload, but they remain on because of their minimum on time constraint.

The total energy consumption of data center modules, in this case, is 221.41 MWh, that is increased by 19.27 MWh compared to Case 1. However, the operation cost of the data centers decreases from $4500 in Case 1 to $2565.4 in this case. The total operation cost of the distribution network is $20,204, which decreased by $1,051 compared to Case 1.

### 6.3.3 Case 3- Coordinated Operation of Data Centers and Distribution Network: Stochastic Solution

In this case, the uncertainties in the demand, renewable generation, requested workloads of the data centers, as well as the volatility in the price of electricity are considered. Using scenario reduction techniques 13 scenarios with respective probabilities were considered.

Table 6-3 represents the operation costs and the associated risks considering multiple values for the trade-off parameter ω. The trade-off parameter captures the balance between the expected operation cost and the cost variability measured by CVaR. The confidence level considered as 95%, ($\alpha = 0.95$). Therefore, the VaR is the lowest operation cost that the probability of obtaining total operation cost higher than that is less than 5%. Calculating the operation cost by setting ω near zero leads to minimum expected operation cost and maximum CVaR. Here, DSO's behavior is risk prone aiming to achieve minimum expected operation cost, ignoring the value of CVaR. Comparing stochastic solution without CVaR i.e. $\omega \approx 0$ with the deterministic solution, it is shown that the operation cost of the data center and the distribution network are increased by 9.2%, and 4%, in stochastic solution respectively.

Table 6-3 Operation cost with associated risk measures

| ω | Data Center Op. Cost | DSO Op. Cost | Total Cost ($) | VaR ($) | CVaR ($) |
|---|---|---|---|---|---|
| ≈ 0 | 2,802 | 21,018 | 23,820 | 25,942 | 26,438 |
| 0.1 | 2,802 | 21,209 | 24,011 | 25,920 | 26,152 |
| 0.2 | 2,766 | 21,436 | 24,202 | 25,918 | 26,110 |
| 0.3 | 2,766 | 21,623 | 24,389 | 25,882 | 26,080 |
| 0.4 | 2,766 | 21,810 | 24,576 | 25,882 | 26,003 |
| 0.5 | 2,766 | 21,998 | 24,764 | 25,836 | 25,936 |
| 0.6 | 2,766 | 22,185 | 24,951 | 25,812 | 25,923 |
| 0.7 | 2,725 | 22,412 | 25,137 | 25,734 | 25,809 |
| 0.8 | 2,725 | 22,591 | 25,316 | 25,721 | 25,792 |
| 0.9 | 2,725 | 22,770 | 25,495 | 25,720 | 25,790 |
| ≈ 1 | 2,725 | 22,947 | 25,672 | 25,720 | 25,790 |

Increasing $\omega$ from 0 to 1 will lead to the increase in the expected operation cost from $23,820, to $25,672. This indicates that increasing the risk aversion would lead to higher expected operation cost, and lower operation cost variability measured by CVaR. As highlighted in Table 6-3, for $\omega > 0.7$ the operation cost increases while the VaR and CVaR remain almost fixed. Consequently, the trade-off between the expected operation cost and risk chosen by DSO is 0.7, as beyond this threshold the expected operation cost increases without an excessive decrease in the operation cost variability.

## 6.4 Summary

The uncertainty of renewable generation resources in distribution network along with the stochasticity of the electricity price affect the demand and operation cost of the distribution networks. This chapter proposes an operation framework to leverage the flexibility of data centers as large electric loads to minimize the operation cost of the distribution network. The proposed framework measures the risk associated with the decisions made by the DSO, through CVaR assessment. The proposed operation framework enables the DSO to avoid over-conservative decisions by setting the trade-off parameter. As a result of the coordinated operation of data centers and distribution network, the total expected operation cost of distribution network decreases and by developing the proposed framework, the risk associated with decisions made is further minimized. It is shown that increasing the conservativeness will increase the expected operation cost and reduces the CVaR.

## 6.5 References


[6.1] N. Nikmehr, S. Najafi-Ravadanegh, "Optimal operation of distributed generations in micro-grids under uncertainties in load and renewable power generation using heuristic algorithm," *IET Renewable Power Generation,* vol. 9, pp. 982-990, 2015.

[6.2] J. Wen, X. N. Han, J. M. Li, Y. Chen, H. Q. Yi, and C. Lu, “Transmission network expansion planning considering uncertainties in loads and renewable energy resources,” *CSEE Journal of Power and Energy Systems,* vol. 1, no. 1, pp. 78–85, Mar. 2015.

[6.3] N. Gast, D. Tomozei, and J. Le Boudec, "Optimal Generation and Storage Scheduling in the Presence of Renewable Forecast Uncertainties," *IEEE Trans. Smart Grid,* vol. 5, pp. 1328-1339, May, 2014.

[6.4] E. Hajipour, M. Bozorg and M. Fotuhi, “Stochastic capacity expansion planning of remote microgrids with wind farms and energy storage,” *IEEE Trans. Sustainable Energy*, vol. 6, pp. 491-498, 2015.

[6.5] EPA report to congress on server and data center energy efficiency. 2007.

[6.6] W. Deng, F. Liu, H. Jin, C. Wu, and X. Liu. Multigreen: cost minimizing multi-source datacenter power supply with online control. *In Proceedings of the fourth international conference on Future energy systems*, pages 149–160. ACM, 2013.

[6.7] R. T. Rockafellar and S. Uryasev, "Optimization of conditional valueat-risk," *J. Risk*, vol. 2, pp. 493–517, Apr. 2000.

[6.8] R. T. Rockefeller and S. Uryasev, "Conditional value-at-risk for general loss distributions," J. *Bank. Financ*., vol. 26, no. 7, pp. 1443–1471, Jul. 2002.

[6.9] N. Noyan, "Risk-averse two-stage stochastic programming with an application to disaster management," *Comput. Oper. Res.*, vol. 39, no. 3, pp. 541–559, Mar. 2012.

[6.10] ieso power data, Available: http://www.ieso.ca/Pages/Power-Data/default.aspx

# Chapter 7

# SUMMARY

Cloud computing provides unique opportunities for the various sectors of the economy such as the automotive industry, education, finance, and governmental entities. In-house IT infrastructures vary in quality and can often be outdated and ineffective. As an alternative, cloud computing is a popular IT model which allows various businesses to "outsource" the IT infrastructure to a more efficient model.

Businesses are opting to use cloud computing in order to eliminate the capital expenses of traditional IT systems and reduce the around-the-clock expert manpower required to operate on-site IT infrastructures. Relying on cloud services gives businesses the flexibility to access a large amount of computing resources easily with minimal hardware and labor.

Data centers as backbone of the cloud computing consist of different physical components such as servers, routers, switches, as well as electricity and cooling infrastructure. Data centers are the brick and mortar face of the cloud computing industry and their performance ensures the prosperity of the underlying business.

Electricity is an essential resource to operate data centers and it is often the costliest expenditure. Furthermore, the power system's reliability and security of energy supply directly translates to data center's operation cost and quality of cloud services. Ironically, the data centers' remarkable demand for electricity influences the power system's reliability and security and on larger scale may affect the local prices of electricity where data centers are installed. This thesis highlights the interdependence between the cloud computing and the power system in planning and operation platforms and addresses the challenges and potential opportunities may infer for both parties in case of cooperation between cloud providers and power system operators.

To address the planning platform, first a coordinated expansion planning for data centers presented which is focused on the economic aspects of the expansion planning of data centers in the data and electricity

networks while ensuring the energy supply security. The presented coordination between the electricity and data networks ensures the adequacy and security of energy supply for the data centers.

In the planning platform, a framework for expansion planning of battery energy storages in the distribution network proposed that captures the interaction among the data center operators and distribution system operators. The proposed framework minimizes the installation and operation cost of the distribution network while ensuring the security and reliability of the distribution network as well as the quality of service for the end users in the data center. As a result of coordination among data center operators and the distribution network operator, the expansion planning costs of the battery energy storage facilities would minimize in the distribution network.

In the operation platform on broad scale, the interaction among the cloud providers in the could computing market as well as the interactions among the cloud providers as demand entities in the wholesale market addressed. As a result, the strategic behavior of the market participants (i.e., cloud providers) in the cloud computing market; as well as the market participants (i.e., cloud providers and generation companies) in the wholesale electricity market to maximize their payoffs presented.

In operation platform on distribution level, the operation of distribution networks with volatile supply and controllable demand of data centers investigated. As a result an operation framework proposed for distribution system operator to determine the demand of distribution network in the day ahead market based on the forecasted day-ahead hourly and intra-hourly electricity prices. The proposed framework incorporates risk aversion by constraining the volatility of the expected operation cost through conditional value at risk (CVaR) assessment.

# APPENDIX

The research conducted to address the problems discussed in this dissertation lead solution methodology to overcome a wide-range of challenges in grid modernization. The relevant publications are listed as following:

This thesis highlights the interdependence between the cloud computing and the power system in planning and operation platforms

- A. Vafamehr and M.E. Khodayar. Network Constrained Expansion Planning of Battery Energy Storages in Distribution Network with Data Centers. IEEE Transaction to be submitted.
- A. Vafamehr, R. Moslemi, and R.K. Sharma Aggregation of BTM Battery Storages to Provide Ancillary Services in Wholesale Markets. Submitted to SEGE, Oshawa, Canada.
- M.E. Khodayar, M.R. Feizi, and A. Vafamehr. Solar Photovoltaic Generation: Benefits and Operation Challenges In Distribution Networks. The Electricity Journal, Elsevier, in press, 2019.
- A. Vafamehr, M. E. Khodayar and K. Abdelghany, "Oligopolistic Competition Among Cloud Providers in Electricity and Data Networks," in IEEE Transactions on Smart Grid, vol. 10, no. 2, pp. 1801-1812, March 2019.
- A. Vafamehr, M. E. Khodayar, S. D. Manshadi, I. Ahmad and J. Lin, "A Framework for Expansion Planning of Data Centers in Electricity and Data Networks Under Uncertainty," in IEEE Transactions on Smart Grid, vol. 10, no. 1, pp. 305-316, Jan. 2019.
- A. Vafamehr and M. E. Khodayar. Energy-aware cloud computing. The Electricity Journal, Elsevier, 31(2): 40-49 March 2018.
- A. Vafamehr and M. E. Khodayar. Operation of Distribution Networks with Volatile Supply and Controllable Data Center Demand. Proceedings of IEEE PES T& D Conference and Exposition, Denver, CO, Apr. 2018.
- M.E. Khodayar, S.D. Manshadi, and A. Vafamehr.The short-term operation of microgrids in a transactive energy architecture. The Electricity Journal, Elsevier, 29(10): 41-48 Dec. 2016.

- A. Vafamehr and M. E. Khodayar. Expansion planning of data centers in energy and cyber networks. Proceedings of IEEE Power and Energy Society General Meeting, Boston, MA , Jul. 2016.