

## Chapter 2

# Energy Cost Minimization for Internet Data Centers Considering Power Outages

**Abstract** With the adoption of smart grid, some cyber-related vulnerabilities may also be introduced. When cyber attacks are launched, the power grid may become unstable and ultimately power outages may occur. In this situation, backup generators would be scheduled to support the operation of Internet data centers. Since the generation cost of backup generators are far higher than the average price of electricity in main grids, higher energy cost for Internet data center operators would be incurred. When power outages caused by cyber attacks occur frequently, the increased energy cost of Internet data center operators would be very large. Thus, it is necessary to consider the operation of Internet data centers in power outage environment. Since Internet data center operators can reduce energy cost by fully utilizing the spatial and temporal diversities of renewable energy in smart microgrids when there are power outages, we consider a scenario that running Internet data centers in smart microgrids and propose an efficient algorithm to minimize the long-term energy cost.

**Keywords** Smart microgrid · Internet data centers · Energy cost · Energy storage · Power outages

## 2.1 Introduction

The past decade has witnessed tremendous growth of online applications and services. Together with the recent trend of cloud computing, massive Internet data centers (IDCs) are deployed for reliability, management, and cost benefits. For IDC operations, a critical issue is the energy consumption. According to a recent study, many IDC operators (e.g., Microsoft and Google) spend more than \$30 million on their annual electricity costs, which contribute to a large portion of IDC operational expenditure [1, 2].

On the other hand, electrical power system is evolving toward smart grid, which has come to describe a next-generation electrical power system that is typified by the increased use of communications and information technology in the generation, delivery, and consumption of electrical energy [3]. Due to its efficiency and potential, lots of research has been done to run distributed IDCs in smart grid environment [4–6], and the benefits are manifold: (1) smart grid technologies (e.g., distributed

generation and distributed electricity storage) can provide energy efficiency savings in IDCs [7]; (2) by accessing smart grid's real-time states (e.g., power price, power demand, power availability), the spatial and temporal variations of electricity price and renewable energy could be exploited to reduce energy cost [8].

With the introduction of smart grid, some cyber-related vulnerabilities may also be created if no appropriate security control is deployed [9]. Therefore, when cyber attacks are launched, power grid may become unstable due to the cyber-power interdependencies in smart grid [10], which may ultimately lead to power outages [11–13]. When power outages occur in main grids to which distributed IDCs are connected, the energy cost of distributed IDCs would increase. The reason is that: (1) fewer spatial and temporal diversities of electricity price could be utilized; (2) backup generators (typically, diesel generators are adopted) in IDCs have to run for a long time (usually several hours [12]) when local renewable energy is in shortage or uninterruptible power system (UPS) capacity is limited. Since the generation cost of backup generators is far higher than the average real-time electricity prices [14–16], higher energy cost would be incurred. When power outages caused by cyber attacks occur frequently, the increased energy cost of IDC operators would be very large. Therefore, it is necessary to consider power outages when running the distributed IDCs in smart grid environment.

To reduce energy cost caused by power outages, it is possible to operate distributed IDCs in smart microgrids (SMGs, which are evolved from microgrids in smart grid environment [17]). The reason is that IDC operators can still exploit the spatial and temporal diversities of renewable energy to reduce energy cost, even if there are power outages in main grids. Specifically, by scheduling workload to the SMGs with abundant renewable energy, the dependence on backup generators in the SMGs with inadequate renewable energy could be reduced, resulting in lower energy cost (i.e., spatial diversity is utilized). On the other hand, energy storage devices can be scheduled to store excess renewable energy and to discharge energy when IDC demands are high, leading to reduced reliance on backup generators and lower energy cost (i.e., temporal diversity is utilized).

As motivated, we study the problem of minimizing the energy cost of distributed IDCs in SMGs considering power outages. Specifically, we consider an IDC operator having some IDCs geographically distributed in several of its own SMGs, which are located in independent electric regions (ERs). Moreover, the IDC operator needs to meet the power demand of IDCs using backup generators, renewable energy sources, and energy storage devices in SMGs and the power purchased from main grids in case of no power outages. The objective of the IDC operator is to minimize the long-term energy cost by deciding the service request distribution, the number of active servers, the operation of energy storage devices, the schedule of backup and renewable generators, as well as the power transactions between SMGs and main grids.

To achieve the above target, the challenge is how to effectively manage the operation of energy storage devices in SMGs [18–20]. Since energy storage devices can bring the time coupling effects to the system, the formulated optimization problem is very difficult to solve. In existing work on data centers [21, 22], several schemes have been developed for exploiting energy storage devices based on Lyapunov

optimization technique [23], which can operate without any statistical information about future uncertain parameters. However, the existing works do not consider running IDCs in SMGs and power outage environment. Thus, new design and analysis are required for an operation algorithm.

The main contributions of this chapter are summarized below:

- By taking SMGs and power outage into account, we formulate a stochastic program problem to minimize the long-term energy cost of distributed IDCs, which captures the service request distribution, server provisioning, battery management, generator scheduling, power transactions between SMGs and main grids.
- We design an operation algorithm to solve the problem based on Lyapunov optimization technique by jointly considering the use of renewable and backup generators, battery management, and electricity purchasing/selling. Moreover, the proposed algorithm can operate without requiring any statistical information about system dynamics.
- We provide the performance guarantee of the proposed algorithm when future random parameters are independent and identically distributed (i.i.d.) over slots. Theoretical analysis shows that the proposed algorithm enables an explicit tradeoff between energy cost saving and battery investment cost.
- Extensive evaluations based on real-world data show the effectiveness of the proposed operation algorithm.

The rest of this chapter is organized as follows. Section 2.2 describes the system model. Problem formulation and algorithm design are conducted in Sect. 2.3. Section 2.4 gives the algorithmic performance analysis and simulations. Finally, conclusions are made in Sect. 2.5.

## 2.2 System Model

We consider an IDC operator with some SMGs located in some independent ERs as shown in Fig. 2.1, where these SMGs are connected to main grids in corresponding ERs. IDCs act as the load in SMGs, while a front-end server acts as a load balancer that receives incoming service requests and dispatches them to IDCs for processing. The architecture of SMG under study is given in Fig. 2.2, where several parts could be highlighted, i.e., generators, loads, storage banks, and energy management system. Generators include backup generators (typically diesel generators) and renewable generators (e.g., solar panels or wind turbines). In this chapter, diesel generators and wind turbines are considered. Diesel generators are fast-responding and dispatchable, and they are typically incorporated in data centers for reliability considerations [2, 24]. By contrast, wind turbines are nondispatchable and their generation outputs depend on weather conditions. Loads in Fig. 2.2 denote IDCs, while storage bank denotes battery. SMGs have two operation modes, i.e., grid-connected and islanded. When there are power outages in main grids, SMGs can isolate themselves from main grids and supply their IDCs using energy storage devices, renewable energy

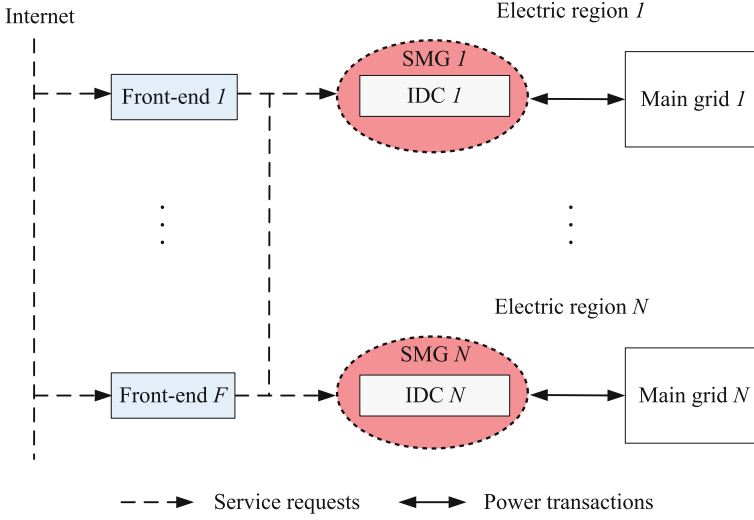


Fig. 2.1 System model

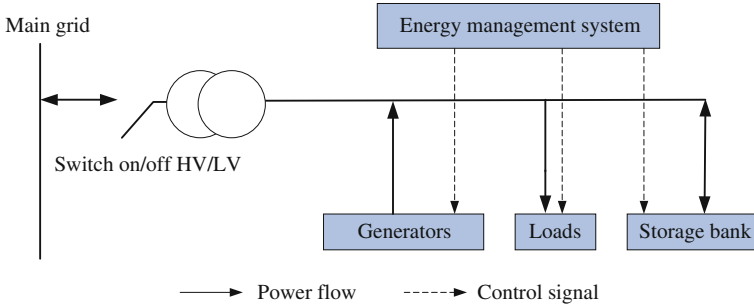


Fig. 2.2 SMG architecture under study

sources as well as backup generators. In the grid-connected mode, power transactions between SMGs and main grids can be conducted.

### 2.2.1 Models Related to Front-End Servers and Internet Data Centers

#### 2.2.1.1 Workload Allocation Model

We consider an IDC operator having  $N$  IDCs geographically distributed in  $N$  independent ERs, and each IDC acts as the load in an SMG connected to a main grid, i.e.,  $N$  SMGs and  $N$  main grids are considered. For simplicity, we use the common

index  $i$  for IDCs, SMGs, and main grids, where  $1 \leq i \leq N$ . In this chapter, we mainly focus on delay-sensitive workloads (e.g., “request-response” type web services [5]). Suppose IDC  $i$  have  $M_i$  homogeneous servers and  $m_i(t)$  of them are turned on to process service requests at slot  $t$ . Define  $\mathbf{m}(t) = (m_1(t), \dots, m_N(t))$  as the active server vector. Moreover,  $m_i(t)$  is assumed to be unchanged at slot  $t$  due to the reliability considerations [25]. We denote the arrival rate of requests at front-end server  $f$  ( $1 \leq f \leq F$ ) as  $R_f(t)$ . Let  $\mathbf{R}(t) = (R_1(t), \dots, R_F(t))$  be the workload arrival vector, where  $F$  is the total number of front-end servers. Let  $\lambda_{f,i}(t)$  be the workload that is assigned from front-end  $f$  to the servers at IDC  $i$  at slot  $t$ .  $\boldsymbol{\lambda}(t) = (\lambda_{f,i}(t), \forall f, i)$  is denoted as the service request distribution. In order to assure that all service requests would be handled, we have

$$\sum_{i=1}^N \lambda_{f,i}(t) = R_f(t). \quad (2.1)$$

### 2.2.1.2 Service Delay Model

To satisfy the quality of service requirements, the average response delay for incoming service requests should be limited within a certain range that is specified in *service level agreement* (SLA).<sup>1</sup> Otherwise, penalties would be incurred. In this chapter, the M/M/n queueing model is adopted to process the incoming workload as in the previous work [4]. Note that the queueing model is not necessarily the most accurate for the practical workload, but it will not affect the nature of the energy cost minimization problem and the proposed operation algorithm. Adopting some general queueing models (e.g., G/G/n as in [26]) would be considered in future work. Let  $D_i^{\max}$  be the threshold that identifies the revenue/penalty region at IDC  $i$ ,  $\mu_i$  be the average service rate of servers in IDC  $i$ . To avoid penalty, the average response delay constraints should be satisfied, i.e.,

$$\frac{1}{m_i(t)\mu_i - \sum_{f=1}^F \lambda_{f,i}(t)} + \frac{1}{\mu_i} \leq D_i^{\max}, \quad (2.2)$$

$$0 \leq m_i(t) \leq M_i. \quad (2.3)$$

### 2.2.1.3 Power Consumption Model

The energy efficiency of an IDC is always measured by *power usage effectiveness* (PUE), which is defined as the ratio of the total power consumption at an IDC to the power consumption at IT equipment. Currently, the typical value of PUE is 2 [1]. Let  $PUE_i$  be the PUE of IDC  $i$ ,  $P_i^{\text{idle}}$  and  $P_i^{\text{peak}}$  represent the idle power and peak power of a server in IDC  $i$ , respectively. Let  $U_i(t)$  be the average server utilization at

---

<sup>1</sup> Here, a simple SLA is adopted as in [26]. Other more complicated SLAs could also be incorporated.

slot  $t$ , which is equal to  $\lambda_i(t)/(m_i(t)\mu_i)$ . Denote the total power consumption of IDC  $i$  at slot  $t$  as  $P_i(t)$  and  $\mathbf{P}(t) = (P_1(t), \dots, P_N(t))$  as the power consumption vector. Following the models in [1],  $P_i(t)$  can be estimated by

$$P_i(t) = m_i(t)(\varphi_i U_i(t) + \gamma_i), \quad (2.4)$$

where  $\varphi_i \triangleq P_i^{\text{peak}} - P_i^{\text{idle}}$ ,  $\gamma_i \triangleq P_i^{\text{idle}} + (\text{PUE}_i - 1)P_i^{\text{peak}}$ . Note that servers in each IDC are assumed to be homogeneous in the above model. For more general situations, the corresponding model can also be incorporated easily.

## 2.2.2 Models Related to Smart Microgrids

### 2.2.2.1 Power Supply Model

Suppose that each SMG has a wind farm comprising several identical wind turbines. Then, the total output of the wind farm is the summation of power outputs at different points of the spatial field. Let  $G_i^r(t)$  be the available wind energy in SMG  $i$  at slot  $t$ , which can be approximated by the model in [27]. Let  $\mathbf{G}^r(t) = (G_1^r(t), \dots, G_i^r(t), \dots, G_N^r(t))$  be the vector of available wind energy at slot  $t$ .

As in [28, 29], the piecewise linear production cost model is adopted to model the energy cost of diesel generators in SMGs. Suppose piecewise linear production cost curve has  $N_B$  segments, each segment  $b$  ( $1 \leq b \leq N_B$ ) introduces a binary variable  $Z_{i,j,b}^g(t)$  and a continuous variable  $Q_{i,j,b}(t)$ , where  $1 \leq j \leq N_i^d$ ,  $N_i^d$  denotes the total number of diesel generators in SMG  $i$ . If the energy produced by the diesel generator  $j$  in SMG  $i$  at slot  $t$  falls into the range  $[g_{i,j,b}^{\min}, g_{i,j,b}^{\max}]$ ,  $Z_{i,j,b}^g(t) = 1$ ; Otherwise,  $Z_{i,j,b}^g(t) = 0$ .  $Q_{i,j,b}(t)$  denotes the specific energy quantity when the energy generation of the diesel generator  $j$  in SMG  $i$  at slot  $t$  fall into the range above. Then, we have

$$Z_{i,j,b}^g(t) \in \{0, 1\}, \quad (2.5)$$

$$\sum_{b=1}^{N_B} Z_{i,j,b}^g(t) \leq 1, \quad (2.6)$$

$$g_{i,j,b}^{\min} Z_{i,j,b}^g(t) \leq Q_{i,j,b}(t) \leq g_{i,j,b}^{\max} Z_{i,j,b}^g(t), \quad (2.7)$$

where  $g_{i,j,b}^{\min}$  and  $g_{i,j,b}^{\max}$  are the lower and upper limit of energy generation of the diesel generator  $j$  in SMG  $i$  corresponding to the segment  $b$  of the piecewise linear production cost model, respectively.

Let  $G_i^a(t)$  be the total power produced by diesel generators in SMG  $i$  at slot  $t$ . Then,  $G_i^a(t)$  is given as

$$G_i^a(t) = \sum_{j=1}^{N_i^d} \sum_{b=1}^{N_B} Q_{i,j,b}(t) / \Delta T, \quad (2.8)$$

where  $\Delta T$  is the duration of a slot. Let  $H_{i,j,b}$  and  $\alpha_{i,j,b}$  be the slope and Y-intercept of the segment  $b$  associated with the diesel generator  $j$  in SMG  $i$ . Then, the total energy cost of diesel generators in SMG  $i$  during time slot  $t$  is given by<sup>2</sup>

$$K_i(t) = \sum_{j=1}^{N_i^d} \sum_{b=1}^{N_B} (H_{i,j,b} Q_{i,j,b}(t) + \alpha_{i,j,b} Z_{i,j,b}^g(t)). \quad (2.9)$$

### 2.2.2.2 Battery Model

$G_i^{g2}(t)$  and  $G_i^{2g}(t)$  represent the power purchased from and sold back to the main grid  $i$  at slot  $t$ , respectively.  $G_i^{b2}(t)$  and  $G_i^{2b}(t)$  are the discharging and charging power for the battery in SMG  $i$  at slot  $t$ . To balance the power supply and demand in SMG  $i$ , we have

$$G_i^a(t) + G_i^r(t) + G_i^{g2}(t) + G_i^{b2}(t) = P_i(t) + G_i^{2b}(t) + G_i^{2g}(t). \quad (2.10)$$

To indicate whether SMG  $i$  is buying or selling power or not at slot  $t$ , two binary variables are adopted, i.e.,  $Z_i^{g2}(t)$  and  $Z_i^{2g}(t)$ . Specifically,  $Z_i^{g2}(t) = 1$  if SMG  $i$  is purchasing power from main grid  $i$  at slot  $t$ ; Otherwise,  $Z_i^{g2}(t) = 0$ . Similarly,  $Z_i^{2g}(t) = 1$  if SMG  $i$  is selling power back to main grid  $i$  at slot  $t$ ; Otherwise,  $Z_i^{2g}(t) = 0$ . Since it is not reasonable to purchase and sell energy on the market at the same time, we have

$$Z_i^{g2}(t), Z_i^{2g}(t) \in \{0, 1\}, \quad (2.11)$$

$$Z_i^{g2}(t) + Z_i^{2g}(t) \leq 1, \quad (2.12)$$

$$0 \leq G_i^{g2}(t) \leq G_i^{b\max} Z_i^{g2}(t), \quad (2.13)$$

$$0 \leq G_i^{2g}(t) \leq G_i^{s\max} Z_i^{2g}(t), \quad (2.14)$$

where  $G_i^{b\max}$  and  $G_i^{s\max}$  denote the maximum purchasing power from main grid  $i$  and maximum selling power to main grid  $i$ , respectively.

Let  $E_i(t)$  be the energy level of the battery in SMG  $i$  at slot  $t$ . Then,  $E_i(t)$  is bounded by the following constraints,

$$E_i^{\min} \leq E_i(t) \leq E_i^{\max}, \quad (2.15)$$

---

<sup>2</sup> For diesel generators, its minimum on/off periods could be regarded as zero and ramping-up/-down rate could be assumed to be  $\infty$  [30]. Thus, some constraints about minimum on/off periods and ramping-up/-down rate are neglected. In addition, due to the lack of public knowledge about startup cost of diesel generators, we also neglect such cost.

where  $E_i^{\max} \geq 0$  is the maximum capacity, and  $E_i^{\min} \geq 0$  is the minimum capacity, which may be set by the battery deep discharge protection settings. As in [21], we do not explicitly consider some practical issues, such as energy leakage in the battery or DC/AC conversion loss, which can be readily incorporated. Then,  $E_i(t)$  has the following dynamics at SMG  $i$ :

$$E_i(t+1) = E_i(t) + \Delta T(G_i^{2b}(t) - G_i^{b2}(t)). \quad (2.16)$$

To indicate whether the battery in SMG  $i$  is charging or not at slot  $t$ , two binary variables are introduced, i.e.,  $Z_i^{bc}(t)$  and  $Z_i^{bd}(t)$ . Specifically,  $Z_i^{bc}(t) = 1$  if that battery is charging; Otherwise,  $Z_i^{bc}(t) = 0$ . Similarly,  $Z_i^{bd}(t) = 1$  if that battery is discharging; Otherwise,  $Z_i^{bd}(t) = 0$ . Without loss of generality, we assume that charging and discharging cannot be done simultaneously [2]. Then, we have

$$Z_i^{bc}(t), Z_i^{bd}(t) \in \{0, 1\}, \quad (2.17)$$

$$Z_i^{bc}(t) + Z_i^{bd}(t) \leq 1, \quad (2.18)$$

$$0 \leq G_i^{2b}(t) \leq G_i^{\text{cm}} Z_i^{bc}(t), \quad (2.19)$$

$$0 \leq G_i^{b2}(t) \leq G_i^{\text{dm}} Z_i^{bd}(t), \quad (2.20)$$

where  $G_i^{\text{cm}}$  and  $G_i^{\text{dm}}$  are the maximum charging and discharging power for the battery in SMG  $i$ , respectively.

### 2.2.3 Models Related to Main Grids

#### 2.2.3.1 Power Outage Model

Power outages could be planned or unplanned. Reasons for unplanned power outages are manifold, for example the faults of transformers and transmission lines caused by environmental or maintenance issues [31], the failures of protective relaying and SCADA caused by cyber attacks [11, 12]. In this chapter, we mainly consider the power outages caused by cyber attacks. Note that the naturally occurring outages could also be incorporated [13]. To analyze the impacts of cyber attacks, RICA approach had been proposed in [12], where power outages caused by cyber attacks are mainly measured by two parameters, i.e., the interval between successful attacks and the time needed to recover from outages, which can be modeled using an exponential-distributed random variable and selected mean time to attack (MTTA) and mean time to recover (MTTR). Let  $Z_i^p(t)$  be an indicator variable that takes the value one whenever there is no power outage in main grid  $i$  at slot  $t$ , and takes the value zero otherwise. Let  $\mathbf{Z}^p(t) = (Z_1^p(t), \dots, Z_N^p(t))$  be the power outage vector. Since power transactions between SMG  $i$  and main grid  $i$  are not allowed when there is power outage, we have

$$0 \leq Z_i^{g2}(t) + Z_i^{2g}(t) \leq Z_i^p(t). \quad (2.21)$$

### 2.2.4 Energy Cost Model

Let  $C(t)$  be the total energy cost of IDC operators at slot  $t$ . Then,  $C(t)$  is composed of several parts, i.e., the energy cost (revenue) associated with buying (selling) electricity, the cost caused by the negative effect of the charging and discharging power [21], and the generation cost of diesel generators in SMGs. That is,  $C(t)$  is given by

$$C(t) = \sum_{i=1}^N \{G_i^{g2}(t)S_i(t) - G_i^{2g}(t)W_i(t)\} \Delta T + \sum_{i=1}^N K_i(t) + \sum_{i=1}^N (Z_i^{bc}(t) + Z_i^{bd}(t)) \rho_i^b, \quad (2.22)$$

where  $\rho_i^b$  denotes the cost incurred by battery charging and discharging per time, which are assumed to be the same [2, 21].  $S_i(t)$  and  $W_i(t)$  denote the price for purchasing and selling electricity from and to main grid  $i$  at slot  $t$ , respectively. Let  $\mathbf{S}(t) = (S_1(t), \dots, S_N(t))$  and  $\mathbf{W}(t) = (W_1(t), \dots, W_N(t))$  be the purchasing and selling price vector, respectively. We suppose that  $S_i(t) \in [S_i^{\min}, S_i^{\max}]$ ,  $W_i(t) \in [W_i^{\min}, W_i^{\max}]$ , which are announced by the utility markets at the beginning of each slot and remains constant in the slot [32].

### 2.2.5 Some Discussions About Models

**Battery Model:**  $G_i^{\text{bmax}}$  and  $G_i^{\text{smax}}$  are determined by the capacity of transformers and power transmission lines [32]. For simplicity, we assume that the transmission loss is negligible and the capacity of transformers and power transmission lines is large enough [32], i.e.,  $G_i^{\text{bmax}} \geq \max_t \{P_i(t) + G_i^{2b}(t)\}$  and  $G_i^{\text{smax}} \geq \max_t \{G_i^a(t) + G_i^r(t) + G_i^{b2}(t)\}$ .

**Power Outage Model:** (1) In RICA approach [12], RICA attack model assumes certain characteristics about the adversary that may not adequately characterize the entire attacker spectrum. For more attacker types, RICA approach needs to be extended [13]; (2) For simplicity, power grid states (i.e., with/without power outages) are assumed to be constant during a slot, but they can change over slots. The assumption is reasonable since the duration of outages caused by cyber attacks is usually several hours [12], which is far larger than  $\Delta T$  (e.g., 15 min or 1 h).

**Energy Cost Model:** (1) We ignore all fixed costs associated with wind generation (e.g., capital expenditure and fixed operational and maintenance cost) in the energy cost model since these parts would not affect the optimization results of the formulated problems P2.1–P2.4. Moreover, according to [33], the variable operational and maintenance cost of wind turbines is negligible. Therefore, the generation cost of wind energy can be assumed to be zero; (2) We assume  $S_i(t)$  and  $W_i(t)$  are independent of the amount of energy to be purchased or sold at slot  $t$ . In addition, we suppose  $S_i(t) \geq W_i(t)$  for all  $t$  [32]. That is, SMG cannot make profit by greedily purchasing energy from the market and then selling it back to the market at a higher price simultaneously.

### 2.3 Problem Formulation and Algorithm Design

In this chapter, we are interested in minimizing the time-average expected energy cost (i.e., the expected energy cost averaged over the infinite time horizon). Compared with existing works on energy cost reduction for data centers [2, 21], we formulate the minimization problem (i.e., P2.1) by taking SMGs and power outage into account. Specifically, the formulated problem jointly consider the renewable and backup generators, battery management, electricity purchasing and selling, as well as power outage.

$$(P2.1) \quad \min \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\{C(t)\}, \quad (2.23a)$$

$$\text{s.t. (2.1)–(2.3), (2.5)–(2.7), (2.10)–(2.21),} \quad (2.23b)$$

where the expectation above is with respect to the stationary distribution of  $\{\mathbf{S}(t), \mathbf{W}(t), \mathbf{R}(t), \mathbf{G}^r(t), \mathbf{Z}^p(t)\}$  and possibly random control decisions that are made in reaction to the observed  $\{\mathbf{S}(t), \mathbf{W}(t), \mathbf{R}(t), \mathbf{G}^r(t), \mathbf{Z}^p(t)\}$ . Specifically, these control decisions are  $\mathbf{m}(t)$ ,  $\lambda(t)$ ,  $\mathbf{G}(t)$ ,  $\mathbf{Z}(t)$  and  $\mathbf{Q}(t)$ . The definitions of  $\mathbf{G}(t)$ ,  $\mathbf{Z}(t)$  and  $\mathbf{Q}(t)$  are given as follows:

$$\mathbf{G}(t) = \{G_i^{g2}(t), G_i^{2g}(t), G_i^{2b}(t), G_i^{b2}(t)\}, \quad (2.24)$$

$$\mathbf{Z}(t) = \{Z_i^{bc}(t), Z_i^{bd}(t), Z_{i,j,b}^g(t), Z_i^{g2}(t), Z_i^{2g}(t)\}, \quad (2.25)$$

$$\mathbf{Q}(t) = \{Q_{i,j,b}(t)\}. \quad (2.26)$$

There are two challenges to solve P2.1. First, the future parameters are unknown, including workload, electricity price, renewable generation output, and power outage state. Second, the constraints on the energy level of battery bring the “time coupling” property to P2.1, which means that the current decision can impact the future decision. Previous methods to handle the “time coupling” problem are usually based on dynamic programming, which suffers from the “the curse of dimensionality” problem. In this section, we design an operation algorithm based on Lyapunov optimization technique. Because of the time-coupling constraint (2.16), we first consider a relaxed problem, which fits into the framework of Lyapunov optimization. Then, we design our algorithm base on the insights provided by the relaxed problem.

#### 2.3.1 Relaxed Problem

In this subsection, a relaxed problem of P2.1 is considered. Defining the time-average expectation of charging and discharging power of the battery at IDC  $i$  under any

feasible control policy of P2.1 as follows,

$$\overline{G_i^{2b}} = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\{G_i^{2b}(t)\}, \quad (2.27)$$

$$\overline{G_i^{b2}} = \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\{G_i^{b2}(t)\}. \quad (2.28)$$

Since battery energy levels update according to (2.16), summing over all  $t \in \{0, 1, 2, \dots, T-1\}$ , taking expectation of both sides and dividing both sides with  $T$  and letting  $T \rightarrow \infty$ , we have  $\overline{G_i^{2b}} = \overline{G_i^{b2}}$ . Then, a relaxed problem is given as follows, called P2.2.

$$(P2.2) \quad \min \quad \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\{C(t)\}, \quad (2.29a)$$

$$\text{s.t. (2.1)–(2.3), (2.5)–(2.7), (2.10)–(2.14), (2.17)–(2.21),} \quad (2.29b)$$

$$\overline{G_i^{2b}} = \overline{G_i^{b2}}. \quad (2.29c)$$

Let  $y^*$  and  $y^{\text{rel}}$  be the optimal solution of P2.1 and P2.2, respectively. As mentioned before, any feasible solution to P2.1 is also a feasible solution to P2.2. Therefore, we have  $y^{\text{rel}} \leq y^*$ . Obviously, P2.2 is easier to solve than P2.1 due to the removal of “time coupling” property. From the framework of Lyapunov optimization [23], we have the following Theorem for the solution to P2.2:

**Theorem 1** *If  $\{\mathbf{S}(t), \mathbf{W}(t), \mathbf{R}(t), \mathbf{G}^r(t), \mathbf{Z}^p(t)\}$  are i.i.d. over slots, then there exists a stationary and randomized policy that takes control decisions (i.e.,  $\hat{\mathbf{m}}(t)$ ,  $\hat{\boldsymbol{\lambda}}(t)$ ,  $\hat{\mathbf{G}}(t)$ ,  $\hat{\mathbf{Q}}(t)$  and  $\hat{\mathbf{Z}}(t)$ ) every slot  $t$  as a pure (possibly randomized) function of the observed  $\{\mathbf{S}(t), \mathbf{W}(t), \mathbf{R}(t), \mathbf{G}^r(t), \mathbf{Z}^p(t)\}$  while satisfying the constraints of P2.2 and providing the following guarantees.*

$$\mathbb{E}\{\hat{G}_i^{2b}(t)\} = \mathbb{E}\{\hat{G}_i^{b2}(t)\}, \quad (2.30)$$

$$\mathbb{E}\{\hat{C}(t)\} = y^{\text{rel}}, \quad (2.31)$$

where the expectation operations above are with respect to the stationary distribution of  $\{\mathbf{S}(t), \mathbf{W}(t), \mathbf{R}(t), \mathbf{G}^r(t), \mathbf{Z}^p(t)\}$ , and randomized control decisions.  $\hat{C}(t)$  denotes the value of (2.22) under the stationary and randomized policy.

The proof can be obtained according to *Theorem of Optimality over  $\omega$ -only Policies* in [23]. Thus, it is omitted for brevity. To obtain such a control policy, we need to know the statistical distributions of all combinations of  $\{\mathbf{S}(t), \mathbf{W}(t), \mathbf{R}(t), \mathbf{G}^r(t), \mathbf{Z}^p(t)\}$ , which may be unknown and difficult to obtain. Moreover, the control policy may be infeasible for the original problem P2.1 as it could violate

the constraint (2.15). However, the existence of such a control policy can help us to derive the performance guarantee of the proposed operation algorithm as shown in the part 3 of the Theorem 2.

### 2.3.2 Proposed Operation Algorithm

The key idea of the proposed operation algorithm is described as follows:

- transforming P2.2 into a queue stability problem according to the framework of Lyapunov optimization technique;
- obtaining the *drift-plus-penalty* term according to the theory of Lyapunov optimization technique;
- minimizing the R.H.S. of the upper bound of the *drift-plus-penalty* term.

Note that the decisions of the proposed operation algorithm do not necessarily satisfy the constraints of P2.1, therefore, we should check the feasibility of the proposed operation algorithm, which would be given in Sect. 2.4.

According to the above key idea, first, we introduce battery virtual queues  $X_i(t)$  in order to transform P2.2 into a queue stability problem, where

$$X_i(t) = E_i(t) - E_i^{\min} - V\theta_i^{\max} - \Delta TG_i^{\text{dm}}, \quad (2.32)$$

where  $\theta_i^{\max} = \max\{S_i^{\max}, W_i^{\max}, H_i^{\max}\}$ ,  $H_i^{\max} = \max_{j,b}\{H_{i,j,b}\}$ .  $V \in [0, V^{\max}]$  is a constant for the tradeoff between minimizing expected energy cost and ensuring the stability of virtue queues. The constant  $V^{\max}$  is carefully selected to ensure that the evolution of the battery energy levels satisfies the constraints in (2.15). Continually, we have

$$X_i(t+1) = X_i(t) + \Delta T(G_i^{2b}(t) - G_i^{b2}(t)). \quad (2.33)$$

Then, according to the framework of Lyapunov optimization in [23], P2.2 can be equivalently transformed to a queue stability problem as follows, named P2.3

$$(P2.3) \min \limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\{C(t)\}, \quad (2.34a)$$

$$\text{s.t. (2.1)–(2.3), (2.5)–(2.7), (2.10)–(2.14), (2.17)–(2.21),} \quad (2.34b)$$

$$\text{Queues } X_i(t) \text{ are mean rate stable.} \quad (2.34c)$$

To solve P2.3, we would develop an operation algorithm based on Lyapunov optimization technique. First, we define a Lyapunov function as follows,

$$L(t) \triangleq \frac{1}{2} \sum_{i=1}^N (X_i(t))^2. \quad (2.35)$$

Then, the one-slot conditional Lyapunov drift is given by

$$\Delta(t) = \mathbb{E}\{L(t+1) - L(t) | \mathbf{X}(t)\}, \quad (2.36)$$

where  $\mathbf{X}(t) = (X_1(t), \dots, X_N(t))$ , the expectation is taken with respect to the randomness of workload, electricity price, renewable generation output, and power outage rate, as well as the randomness in choosing the control decisions. We define  $\chi_i(t) = G_i^{2b}(t) - G_i^{b2}(t)$ , then

$$L(t+1) - L(t) = \frac{1}{2} \sum_{i=1}^N [(X_i(t+1))^2 - (X_i(t))^2] \leq \sum_{i=1}^N [\omega_i + X_i(t) \Delta T \chi_i(t)], \quad (2.37)$$

where  $\omega_i \triangleq \frac{(\Delta T \max\{G_i^{\text{cm}}, G_i^{\text{dm}}\})^2}{2}$ . Thus, we have

$$\Delta(t) \leq \sum_{i=1}^N \omega_i + \sum_{i=1}^N \mathbb{E}\{X_i(t) \Delta T \chi_i(t) | \mathbf{X}(t)\}. \quad (2.38)$$

Continually, following the Lyapunov optimization framework, we add a function of the expected energy cost over one period (i.e., the penalty function) to (2.36) to obtain the *drift-plus-penalty* term as follows,

$$\Delta Y(t) = \Delta(t) + V \mathbb{E}\{C(t) | \mathbf{X}(t)\} \leq \sum_{i=1}^N \omega_i + \mathbb{E}\left\{ \sum_{i=1}^N X_i(t) \Delta T \chi_i(t) + VC(t) | \mathbf{X}(t) \right\}. \quad (2.39)$$

Note that the proposed operation algorithm intends to minimize the R.H.S. of the upper bound of the *drift-plus-penalty* term subject to the constraints in P2.3, which is equivalent to minimize P2.4 based on the observed system states at the beginning of slot  $t$ . Consequently, the proposed operation algorithm can be described by Algorithm 1.

$$(\text{P2.4}) \quad \min \quad \sum_{i=1}^N \{X_i(t) \Delta T \chi_i(t)\} + VC(t) \quad (2.40a)$$

$$\text{s.t. (2.1)–(2.3), (2.5)–(2.7), (2.10)–(2.14), (2.17)–(2.21),} \quad (2.40b)$$

---

**Algorithm 1** : Operation Algorithm for IDC Operators

---

- 1: **For** each slot  $t$  **do**
  - 2: At the beginning of slot  $t$ , observe system states:  $X_i(t)$ ,  $W_i(t)$ ,  $R_f(t)$ ,  $S_i(t)$ ,  $G_i^r(t)$  and  $Z_i^p(t)$ ;
  - 3: Choose control decisions  $\mathbf{m}(t)$ ,  $\boldsymbol{\lambda}(t)$ ,  $\mathbf{G}(t)$ ,  $\mathbf{Q}(t)$ , and  $\mathbf{Z}(t)$  as the solution to P2.4;
  - 4: Update  $X_i(t)$  according to (2.33);
  - 5: **End**
-

### 2.3.3 Solution

To solve P2.4, which is a MILP, its certain structure can be exploited, i.e., by fixing the integer variables first, the resulting problem becomes convex for continuous variables. Therefore, Benders' decomposition [34] can be adopted to solve P2.4 in this chapter, which can decompose P2.4 into two problems, i.e., master problem and subproblem. In addition, considering that there are huge number of servers in IDCs and a large fraction of them are active, we can relax the integer constraint on  $m_i(t)$  without significant energy cost penalties [4].

Before giving the algorithm to solve P2.4, we define some problems in the following part.

Let  $\mathbf{x} \triangleq \mathbf{Z}(t)$  and  $\mathbf{z} \triangleq \{\mathbf{m}(t), \boldsymbol{\lambda}(t), \mathbf{G}(t), \mathbf{Q}(t)\}$ . If  $\mathbf{x}$  is fixed to feasible integer configuration  $\bar{\mathbf{x}} \triangleq \bar{\mathbf{Z}}(t)$ , then, P2.4 can be reduced to P2.5 as follows,

$$\begin{aligned}
 \text{(P2.5)} \quad \min_{\mathbf{z}} \quad & \sum_i \{X_i(t) \Delta T (G_i^{2b}(t) - G_i^{2g}(t))\} \\
 & + \sum_i \{V(G_i^{2g}(t) S_i(t) - G_i^{2g}(t) W_i(t)) \Delta T\} \\
 & + \sum_i \sum_j \sum_b V H_{i,j,b} Q_{i,j,b}(t)
 \end{aligned} \tag{2.41a}$$

$$\text{s.t. } \Gamma(\mathbf{z}) \leq \mathbf{0}, \tag{2.41b}$$

where  $\Gamma(\mathbf{z})$  is obtained by fixing the discrete parameters in (2.1)–(2.3), (2.7), (2.10), (2.13), (2.14), (2.19), (2.20) with  $\bar{\mathbf{x}}$ .

Let P2.6 be the dual of the P2.5, and the objective function of P2.6 is denoted as  $\Upsilon(\bar{\mathbf{x}}, \boldsymbol{\phi})$ , the constraints of P2.6 are described by  $\Lambda(\boldsymbol{\phi}) \leq \mathbf{0}$ . Then, P2.6 is given by

$$\text{(P2.6)} \quad \max_{\boldsymbol{\phi}} \quad \Upsilon(\bar{\mathbf{x}}, \boldsymbol{\phi}) \tag{2.42a}$$

$$\text{s.t. } \Lambda(\boldsymbol{\phi}) \leq \mathbf{0}. \tag{2.42b}$$

Continually, the Benders' subproblem can be defined as follows (P2.7, which is a linear programming problem and can be solved in polynomial time),

$$\begin{aligned}
 \text{(P2.7)} \quad \Psi^* = \max_{\boldsymbol{\phi}} \quad & \sum_i V(\overline{Z_i^{bc}(t)} + \overline{Z_i^{bd}(t)}) \rho_i^b + \Upsilon(\bar{\mathbf{x}}, \boldsymbol{\phi}) \\
 & + \sum_i \sum_j \sum_b V \alpha_{i,j,b} \overline{Z_{i,j,b}^g(t)}
 \end{aligned} \tag{2.43a}$$

$$\text{s.t. } \Lambda(\boldsymbol{\phi}) \leq \mathbf{0}. \tag{2.43b}$$

We define the Benders' master problem below (called P2.8, which is a 0–1 integer programming problem and could be solved by a pseudo-polynomial time algorithm)

$$\text{(P2.8)} \quad \Omega^* = \min_{\mathbf{x}} \Omega \tag{2.44a}$$

$$\begin{aligned} \text{s.t. } \Omega \geq & \sum_i V(Z_i^{\text{bc}}(t) + Z_i^{\text{bd}}(t))\rho_i^b + \Upsilon(\mathbf{x}, \phi^{k_2}) \\ & + \sum_i \sum_j \sum_b V\alpha_{i,j,b} Z_{i,j,b}^g(t), \quad 1 \leq k_2 \leq K_2, \end{aligned} \quad (2.44\text{b})$$

$$\Upsilon(\mathbf{x}, \phi^{k_1}) \leq 0, \quad 1 \leq k_1 \leq K_1, \quad (2.44\text{c})$$

$$(2.5), (2.6), (2.11), (2.12), (2.17), (2.18), (2.21), \quad (2.44\text{d})$$

where the Benders' optimality cut (4b) and the Benders' feasibility cut (4c) are incorporated when the Benders' subproblem P2.7 is bounded and unbounded, respectively.  $K_1$  and  $K_2$  are the number of feasibility cut and optimality cut, respectively.  $\phi^{k_1}$  and  $\phi^{k_2}$  are the unbounded rays and the extreme points, respectively, where the method to obtain the unbound rays can be found in [35].

Based on above optimization problems, an iterative algorithm based on Benders' decomposition to solve P2.4 is proposed in Algorithm 1. Since  $K_1$  and  $K_2$  are finite, the iterative algorithm is expected to converge to the optimal solution within a finite number of iterations.

---

**Algorithm 2** : Procedures of solving P2.4

---

```

1: Input: An initial feasible solution  $\mathbf{x}^1$ 
   and a convergence tolerant parameter  $\varepsilon > 0$ 
2: Output: An optimal solution  $\mathbf{x}^*$  and  $z^*$ 
3: {Initialization}
4: Given initial feasible integer solution  $\mathbf{x}^1$ 
5:  $\text{LB}^1 := -\infty$ 
6:  $\text{UB}^1 := \infty$ 
7:  $q=1, K_1=0, K_2=0$ ;
8: for iteration  $q$  do
9:   {solve subproblem P2.7}
10:  if unbounded then
11:    Get unbounded ray  $\phi^{K_1+1}, K_1=K_1+1$ 
12:    Add feasibility cut  $\Upsilon(\mathbf{x}, \phi^{K_1}) \leq 0$  to P2.8
13:  else
14:    Get extreme point  $\phi^{K_2+1}, K_2=K_2+1$ 
15:    Add optimality cut  $\Omega \geq \sum_i V(Z_i^{\text{bc}}(t) + Z_i^{\text{bd}}(t))\rho_i^b + \sum_i \sum_j \sum_b V\alpha_{i,j,b} Z_{i,j,b}^g(t) + \Upsilon(\mathbf{x}, \phi^{K_2})$ 
   to P2.8
16:     $\text{UB}^q := \min\{\text{UB}^{q-1}, \Psi^*\}$ 
17:  end
18:  {solve master problem P2.8}
19:   $\text{LB}^q := \Omega^*$ 
20:  if  $\text{LB}^q + \varepsilon > \text{UB}^q$  then
21:    Obtain the optimal  $\mathbf{x}^*$ 
22:    Solve P2.4 to recover solution by fixing  $\bar{\mathbf{x}}$  with  $\mathbf{x}^*$ 
23:    return optimal solution  $\mathbf{x}^*$  and  $z^*$ 
24:  else
25:     $q = q + 1$ 
26:  end
27: end

```

---

## 2.4 Analysis and Simulations

As explained in Sect. 2.3.2, we need to check the feasibility of the proposed operation algorithm. In this section, we present a lemma about the operation of battery. Then, based on the lemma, we give a theorem about the feasibility of the proposed algorithm, i.e., the proposed algorithm can operate without requiring any statistical knowledge. Moreover, we provide the performance guarantee of the proposed algorithm when  $\{\mathbf{S}(t), \mathbf{W}(t), \mathbf{R}(t), \mathbf{G}^T(t), \mathbf{Z}^P(t)\}$  are i.i.d. over slots. Note that algorithms developed based on Lyapunov optimization technique are also robust to non-i.i.d. and non-ergodic behaviors of the stochastic processes (e.g., [23]). Therefore, performance evaluations in next section are conducted based on practical data without any specific distribution assumption, which shows the robustness of the proposed operation algorithm.

### 2.4.1 Analysis

First, we present a Lemma that is useful for the analysis of algorithmic performance.

**Lemma 1** *Let  $\theta_i^{\min} = \min\{S_i^{\min}, W_i^{\min}, H_i^{\min}\}$ ,  $H_i^{\min} = \min_{j,b}\{H_{i,j,b}\}$ , The optimal solution to P2.4 has the following properties:*

1. *If  $X_i(t) < -V\theta_i^{\max}$ , the optimal solution always choose  $G_i^{b2}(t) = 0$ ,*
2. *If  $X_i(t) > -V\theta_i^{\min}$ , the optimal solution always choose  $G_i^{2b}(t) = 0$ .*

*Proof*

1. For each SMG  $i$ , suppose that  $X_i(t) < -V\theta_i^{\max}$  and  $G_i^{b2}(t) > 0$ , then, we have  $G_i^{2b}(t) = 0$ . Let  $\Gamma_1(t)$  denote the optimal value of the objective in P2.4, which is the summation of  $\Gamma_{1,i}(t)$  at all IDCs. To prove that the above decision is not optimal, we choose  $\tilde{G}_i^{b2}(t) = 0$ , and  $\tilde{G}_i^{2b}(t) = 0$ . Let  $\Gamma_2(t)$  denote the value of the objective in P2.4, which is the summation of  $\Gamma_{2,i}(t)$  at all IDCs. Given the power demand  $P_i(t)$ , the decisions about power supply for IDCs can be divided into the following cases:

**Case 1:** If  $G_i^{g2}(t) = G_i^{2g}(t) = 0$ , we set  $\tilde{G}_i^{g2}(t) = G_i^{g2}(t)$ ,  $\tilde{G}_i^{2g}(t) = G_i^{2g}(t)$ , then,  $G_i^a(t) + G_i^{b2}(t) = \tilde{G}_i^a(t)$ . Continually, we have

$$\Gamma_{1,i}(t) - \Gamma_{2,i}(t) \geq -(X_i(t) + VH_i^{\max})\Delta TG_i^{b2}(t) + V\rho_i^b > 0. \quad (2.45)$$

**Case 2:** If  $G_i^{g2}(t) > 0$ ,  $G_i^{2g}(t) = 0$ , we set  $\tilde{G}_i^{2g}(t) = G_i^{2g}(t)$ ,  $\tilde{G}_i^a(t) = G_i^a(t)$ . Then,  $G_i^{g2}(t) + G_i^{b2}(t) = \tilde{G}_i^{g2}(t)$ . Continually, we have

$$\Gamma_{1,i}(t) - \Gamma_{2,i}(t) \geq -(X_i(t) + VS_i^{\max})\Delta TG_i^{b2}(t) + V\rho_i^b > 0. \quad (2.46)$$

**Case 3:** If  $G_i^{g2}(t) = 0$ ,  $G_i^{2g}(t) > 0$ , we set  $\tilde{G}_i^{g2}(t) = G_i^{g2}(t)$ ,  $\tilde{G}_i^a(t) = G_i^a(t)$ . Then,  $G_i^{2g}(t) - G_i^{b2}(t) = \tilde{G}_i^{2g}(t)$ . Continually, we have

$$\Gamma_{1,i}(t) - \Gamma_{2,i}(t) \geq -(X_i(t) + VW_i^{\max})\Delta TG_i^{b2}(t) + V\rho_i^b > 0. \quad (2.47)$$

Taking three cases into consideration, we know that when  $X_i(t) < -V\theta_i^{\max}$ , the optimal decision always choose  $G_i^{b2}(t) = 0$ .

2. Similar to the part 1 of Lemma 1, part 2 can be proved. Thus, the proof of this part is omitted for brevity.

Then, we analyze the feasibility and performance guarantee of the proposed operation algorithm in the following Theorem.

**Theorem 2** Suppose the initial battery energy level  $E_i(0) \in [E_i^{\min}, E_i^{\max}]$ . Implementing the proposed operation algorithm with any fixed parameter  $V \in (0, V^{\max}]$ , we have the following properties:

1. The battery energy level  $E_i(t)$  is always in the range  $[E_i^{\min}, E_i^{\max}]$  for all slots. Moreover, queues  $X_i(t)$  are mean rate stable.
2. The decisions of the proposed operation algorithm are feasible to P2.1, i.e., the proposed operation algorithm can operate without requiring any statistical knowledge about system dynamics.
3. If  $\{\mathbf{S}(t), \mathbf{W}(t), \mathbf{R}(t), \mathbf{G}^r(t), \mathbf{Z}^p(t)\}$  are i.i.d. over slots and  $\omega = \sum_i \omega_i$ , then the time-average expected energy cost under the proposed operation algorithm is within bound  $\omega/V$  of the optimal value:  $\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\{C(t)\} \leq y^* + \frac{\omega}{V}$ .

*Proof*

1. To show  $E_i(t) \in [E_i^{\min}, E_i^{\max}]$ , according to the definition of  $X_i(t)$ , it is equivalent to show that for each IDC  $i$ ,

$$X_i(t) \geq -V\theta_i^{\max} - \Delta TG_i^{\text{dm}}, \quad (2.48)$$

and

$$X_i(t) \leq E_i^{\max} - E_i^{\min} - V\theta_i^{\max} - \Delta TG_i^{\text{dm}}. \quad (2.49)$$

As  $E_i^{\min} \leq E_i(0) \leq E_i^{\max}$ , the above inequalities hold for  $t = 0$ . We prove that the constraints are satisfied for all periods by induction. Suppose the above inequalities hold for slot  $t$ , we need to prove that they also hold for slot  $t + 1$ .

- If  $-V\theta_i^{\max} - \Delta TG_i^{\text{dm}} \leq X_i(t) < -V\theta_i^{\max}$ , then, according to Lemma 1, the optimal decision would choose  $G_i^{b2}(t) = 0$ . Thus,  $X_i(t+1) \geq X_i(t) \geq -V\theta_i^{\max} - \Delta TG_i^{\text{dm}}$ . If  $-V\theta_i^{\max} \leq X_i(t) < E_i^{\max} - E_i^{\min} - V\theta_i^{\max} - \Delta TG_i^{\text{dm}}$ , then, according to the dynamics of  $X_i(t)$ ,  $X_i(t+1) \geq -V\theta_i^{\max} - \Delta TG_i^{b2}(t) > -V\theta_i^{\max} - \Delta TG_i^{\text{dm}}$ .

- If  $-V\theta_i^{\min} < X_i(t) \leq E_i^{\max} - E_i^{\min} - V\theta_i^{\max} - \Delta T G_i^{\text{dm}}$ , then, According to Lemma 1, the optimal decision would choose  $G_i^{2b}(t) = 0$ . Thus,  $X_i(t+1) \leq X_i(t) \leq E_i^{\max} - E_i^{\min} - V\theta_i^{\max} - \Delta T G_i^{\text{dm}}$ . If  $-V\theta_i^{\max} - \Delta T G_i^{\text{dm}} \leq X_i(t) \leq -V\theta_i^{\min}$ , then,  $X_i(t+1) \leq -V\theta_i^{\min} + \Delta T G_i^{\text{cm}} \leq E_i^{\max} - E_i^{\min} - V\theta_i^{\max} - \Delta T G_i^{\text{dm}}$ , where we have used the following condition,

$$V \leq \frac{E_i^{\max} - E_i^{\min} - \Delta T(G_i^{\text{cm}} + G_i^{\text{dm}})}{\theta_i^{\max} - \theta_i^{\min}}. \quad (2.50)$$

Continually, the upper bound of parameter  $V$  is given by

$$V^{\max} = \min_i \frac{E_i^{\max} - E_i^{\min} - \Delta T(G_i^{\text{cm}} + G_i^{\text{dm}})}{\theta_i^{\max} - \theta_i^{\min}}, \quad (2.51)$$

where  $\theta_i^{\min} = \min\{W_i^{\min}, S_i^{\min}, H_i^{\min}\}$ ,  $H_i^{\min} = \min_{j,b}\{H_{i,j,b}\}$ . Since  $\limsup_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}\{|X_i(T)|\} = 0$ , we can know that queues  $X_i(t)$  are mean rate stable.

2. The proposed operation algorithm can make decisions to satisfy all constraints in P2.4 and update  $X_i(t)$  normally (i.e., update  $E_i(t)$ ). Meanwhile, taking the part 1 of Theorem 2 into consideration, we know that all constraints in P2.1 can be satisfied by the decisions of the proposed algorithm. That is, the proposed algorithm can operate without requiring any statistical knowledge.
3. By plugging the policy mentioned in Theorem 1 into the R.H.S. of *drift-plus-penalty* term, then, we have

$$\Delta Y(t) \leq \omega + V\mathbb{E}\{\hat{C}(t)\} \leq \omega + Vy^*. \quad (2.52)$$

Taking the expectation of both sides, using the law of iterative expectation, we have

$$\mathbb{E}[\Delta Y(t)] = \mathbb{E}[\Delta(t)] + V\mathbb{E}[C(t)] \leq \omega + Vy^*. \quad (2.53)$$

Then, summing the above equations over  $t \in \{0, 1, 2, \dots, T-1\}$ , we have

$$V \sum_{t=0}^{T-1} \mathbb{E}\{C(t)\} \leq \omega T + VTy^* - \mathbb{E}\{L(T)\} + \mathbb{E}\{L(0)\}. \quad (2.54)$$

Dividing both side by  $VT$ , and taking a  $\limsup$  of both sides. Let  $T \rightarrow \infty$ , and using the facts that  $E\{L(0)\}$  is finite and  $E\{L(T)\}$  is nonnegative, we arrive at the following performance guarantee,

$$\limsup_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\{C(t)\} \leq y^* + \frac{\omega}{V}. \quad (2.55)$$

It is worth noting that the choice of  $V$  controls the optimality of the proposed operation algorithm. Specifically, a larger  $V$  leads to a tighter optimality gap. However, according to the definition of  $V^{\max}$ , we know that larger  $V$  requires larger investment on battery capacity, which would result in larger investment cost since batteries are very expensive currently. Therefore, there is a tradeoff between energy cost saving and battery investment cost.

## 2.4.2 Simulations

In this section, we evaluate the performance of the proposed operation algorithm based on real-world traces. Our purpose is twofold: (1) to show the effectiveness of our proposed operation algorithm under power outage environment; (2) to show the tradeoff between energy cost saving and battery investment cost.

### 2.4.2.1 Real-World Traces and Experimental Setup

**System Parameters.** We consider a scenario that the total workload from one front-end web portal server is forwarded to two back-end IDCs located in two independent ERs, i.e.,  $F = 1$ ,  $N = 2$ . Some parameters about servers in IDCs are set as follows:  $P_1^{\text{peak}} = 250 \text{ W}$ ,  $P_2^{\text{peak}} = 240 \text{ W}$ ,  $P_1^{\text{idle}} = 175 \text{ W}$ ,  $P_2^{\text{idle}} = 168 \text{ W}$ ,  $\text{PUE}_1 = 2.5$ ,  $\text{PUE}_2 = 2.4$  [1],  $M_1 = 22,000$ ,  $M_2 = 35,000$ ,  $\mu_1 = 2$  requests/s,  $\mu_2 = 1.75$  requests/s,  $D_i^{\max} = 0.7 \text{ s}$ ,  $\Delta T = 1 \text{ h}$  [2]. In this chapter, we assume that generators in each IDC have the capacity to power the whole IDC, which is quite common in industry. Specifically, each IDC has 10 diesel generators with model TP-C2000-T2-60 (2 MW) and 5 diesel generators with model TP-C100-T1-60 (100 kW) [14], where the parameters  $H_{i,j,b}$  and  $\alpha_{i,j,b}$  can be estimated as in [29, 36]. According to the recent empirical experiments, we assume that the limits of battery charging/discharging rates are  $P^{\text{cm}} = P^{\text{dm}} = 0.5 \text{ MW}$ , and charging/discharging cost is  $\rho_i^b = 0.1$  dollars [21]. In addition, we set  $V = V^{\max}$ .

**Wind Energy Data.** The wind speed information is obtained from the dataset published by the National Renewable Energy Laboratory.<sup>3</sup> Based on the 1 min raw data, we calculate the average wind speed during the disjoint hour. We adopt the V90 wind turbine model in this chapter.<sup>4</sup>

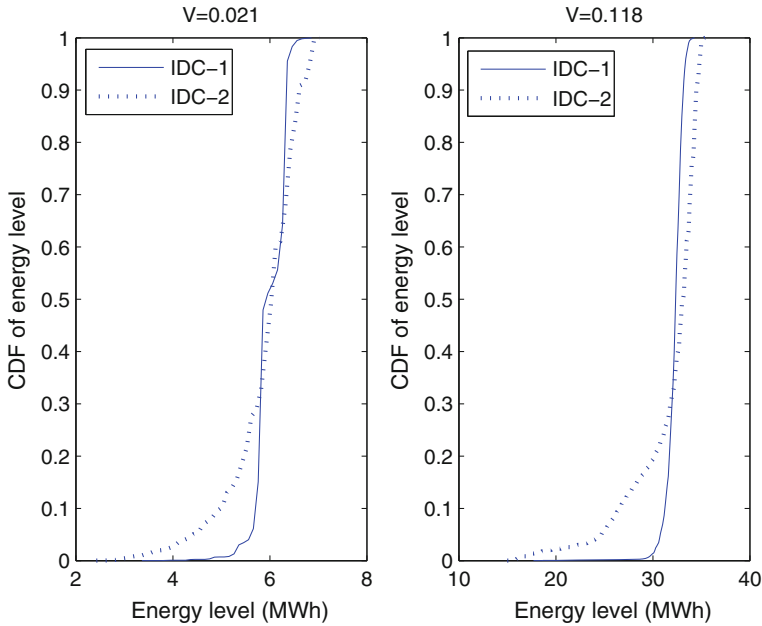
**Electricity Price Data.** The real-time electricity price traces of the U.S. NYISO and ERCOT markets in 2012 are used. Based on the traces, we calculate the average electricity price during the disjoint hour. Moreover, we set  $W_i(t) = 0.9S_i(t)$  [37].

**Workload Data.** The real-world workload data is taken from FIFA's 1998 World Cup web traces between June 10 and June 30, 1998.<sup>5</sup> Considering the increase of

<sup>3</sup> <http://www.nrel.gov/midc>, Sept. 2013.

<sup>4</sup> <http://www.vestas.com/en/wind-power-plants/>, Sept. 2013.

<sup>5</sup> <http://ita.ee.lbl.gov/html/traces.html>, Sept. 2013.



**Fig. 2.3** CDF of energy levels with varying  $V$

Internet traffic in the past decade, we modify the original workload by enlarging 60 times<sup>6</sup> and repeat it to get a 365-day workload trace.

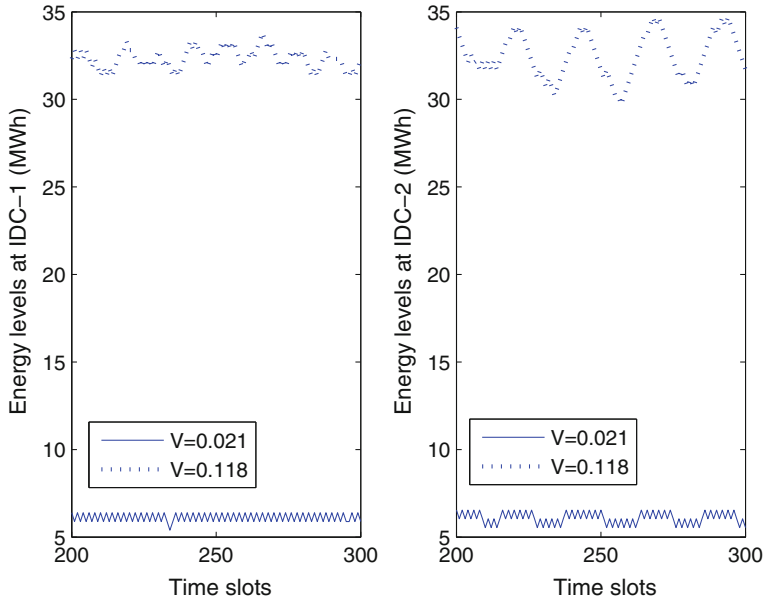
**Power Outage Data.** Power outage process in each ER is modeled using an exponential-distributed random variable and some selected MTTA and MTTR [12]. We set  $MTTA_1 = MTTA_2 = 876$  h,  $MTTR_1 = MTTR_2 = 8$  h, where  $MTTA-i$  and  $MTTR-i$  are the parameters associated with the main grid  $i$ . The method of generating the power outage process can be found in [13].

**Benchmark.** To show the effectiveness of the proposed algorithm, we adopt a benchmark called cost-aware dynamic provisioning (CDP) based on the scheme proposed by Rao et al. [5]. Note that CDP can exploit the spatial diversity of electricity price by traffic routing. However, renewable energy and battery are not considered in [5]. For more fair comparison, renewable energy is adopted in CDP, which intends to minimize the total energy cost in each slot based on the observed system states.

#### 2.4.2.2 Experimental Results

Figure 2.3 illustrates the cumulative distribution function (CDF) of energy levels under the proposed algorithm and varying control parameter  $V$ . It can be observed

<sup>6</sup> E.g., the average workload is smaller than 2 million requests/h, while the average workload of Google search is 121 million requests/h.



**Fig. 2.4** Energy levels at different IDCs

that the proposed algorithm is feasible since the energy levels are fluctuating in their normal ranges. As shown by Fig. 2.4, smaller  $V$  would lead to more frequent discharging or charging process since the proposed algorithm at this time put more weight on maintaining the energy queue stability rather than energy cost minimization.

As depicted in Fig. 2.5, larger  $V$  would lead to lower energy cost under the proposed algorithm, which validates the algorithmic performance in the part 3 of Theorem 2. Though the part 3 of Theorem 2 holds when  $\{\mathbf{S}(t), \mathbf{W}(t), \mathbf{R}(t), \mathbf{G}^T(t), \mathbf{Z}^P(t)\}$  are i.i.d. over slots, the simulation results based on real-world traces (without any specific distribution assumption) show that the proposed algorithm is robust to non-i.i.d. and non-ergodic cases (note that the theoretical basis for such robustness can be found in [23]).

In addition, it can be seen that power outages would lead to the increase of total energy cost under two schemes. Specifically, the increased energy cost is 59,097 dollars and 55,067 dollars under the CDP and the proposed algorithm, respectively. It is expected that the increased energy cost would become larger as MTTA decreases and MTTR increases. Compared with the CDP, the proposed algorithm achieves the better performance and the performance gap between the proposed algorithm and the CDP is larger as  $V$  increases. The reason is given as follows: larger  $V$  means larger battery capacity, which can better utilize the spatial and temporal diversities of electricity price and wind energy. In contrast, CDP can merely exploit the spatial diversities of electricity price and wind energy. The same reason can be used to explain the phenomenon in Fig. 2.6, i.e., the profit under the proposed algorithm is

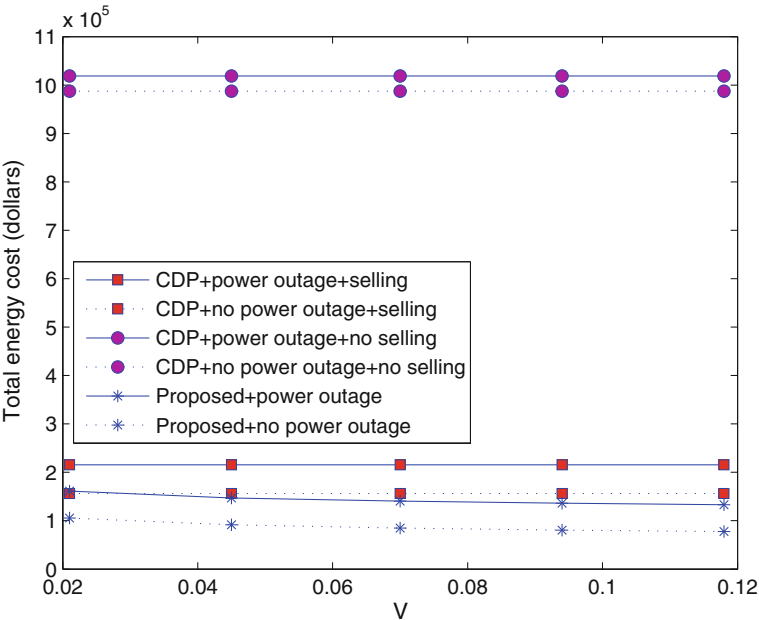


Fig. 2.5 Total energy cost under different algorithms

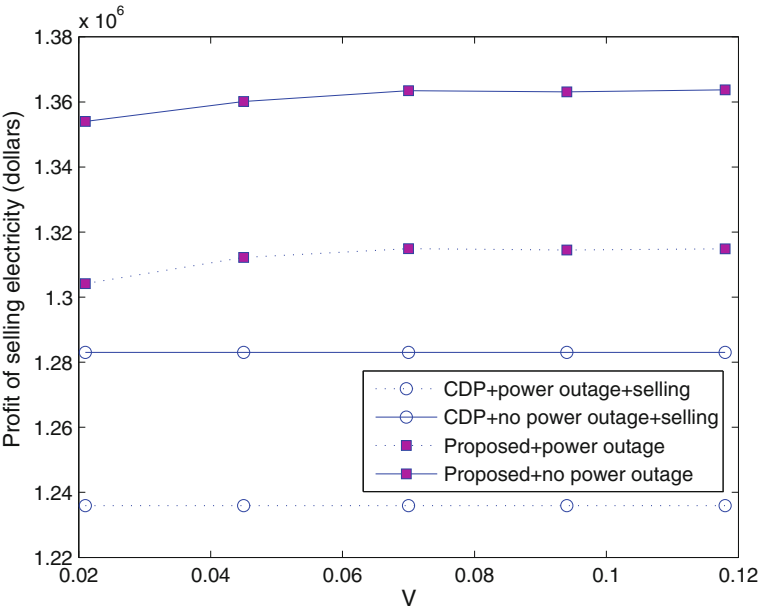


Fig. 2.6 Profit due to the electricity selling

increasing as  $V$  increases. In addition, it is obvious that the profit due to electricity selling is lower when there are power outages since power transactions under the above situation are not allowed.

## 2.5 Summary

In this chapter, we studied the problem of minimizing the long-term energy cost of distributed IDCs in smart microgrids considering power outages. At first, we formulated the problem as a stochastic program by taking SMGs and power outage into consideration. To solve the problem, an operation algorithm was designed based on Lyapunov optimization technique. Moreover, we provided the algorithmic performance guarantee when random parameters are i.i.d. over slots. In addition, the proposed algorithm enables an explicit tradeoff between energy cost saving and battery investment cost. Finally, extensive evaluations based on real-world data showed the effectiveness of the proposed algorithm.

## References

1. Qureshi A, Weber R, Balakrishnan H, Gutttag J, Maggs B (2009) Cutting the electric bill for internet-scale systems. In: Proceedings of ACM special interest group on data communication (SIGCOMM)
2. Guo Y, Fang Y (2013) Electricity cost saving strategy in data centers by using energy storage. *IEEE Trans Parallel Distrib Syst* 24(6):1149–1160
3. <http://smartgrid.ieee.org/ieee-smart-grid>. Accessed 23 Sept 2013
4. Wang P, Rao L, Liu X, Qi Y (2012) D-pro dynamic data center operations with demand-responsive electricity prices in smart grid. *IEEE Trans Smart Grid* 4(3):1–12
5. Rao L, Liu X, Xie L, Liu W (2012) Coordinated energy cost management of distributed internet data centers in smart grid. *IEEE Trans Smart Grid* 3(1):50–58
6. Yu L, Jiang T, Cao Y, Yang S, Wang Z (2012) Risk management in internet data center operations under smart grid environment. In: Proceedings of IEEE international conference on smart grid communications (SmartGridComm)
7. Salomonsson D, Soder L, Sannino A (2008) An adaptive control system for a DC microgrid for data centers. *IEEE Trans Ind Appl* 44(6):2414–2421
8. Rahman A, Liu X, Kong F (2014) A survey on geographic load balancing based data center power management in the smart grid environment. *IEEE Commun Surv Tutor* 16(1):214–233
9. Mo Y, Kim TH-J, Brancik K, Dickinson D, Lee H, Perrig A, Sinopoli B (2011) Cyber-physical security of a smart grid infrastructure. *Proc IEEE* 100(1):195–209
10. Falahati B, Fu Y, Wu L (2012) Reliability assessment of smart grid considering direct cyber-power interdependencies. *IEEE Trans Smart Grid* 3(3):1515–1524
11. Kirschen D, Bouffard F (2009) Keep the lights on and the information flowing, a new framework for analyzing power system security. *IEEE Power Energy Mag* 7(1):50–60
12. Stamp J, McIntyre A, Ricardson B (2009) Reliability impacts from cyber attack on electric power systems. In: Proceedings of power systems conference and exposition (PSCE)
13. Stamp J, Colbaugh R, Laviolette R, McIntyre A, Richardson B (2009) Impacts analysis for cyber attack on electric power systems (National SCADA Test Bed FY08). SAND2009-1673

14. Dufo-López R, Bernal-Agustín JL, Yusta-Loyo JM et al (2011) Multi-objective optimization minimizing cost and life cycle emissions of stand-alone PV-wind-diesel systems with batteries storage. *Appl Energy* 88(11):4033–4041
15. Gasoline and Diesel Fuel Update. Available via DIALOG. <http://www.eia.gov/petroleum/gasdiesel/>. Accessed 23 Sept 2013
16. Federal Energy Regulatory Commission. Available via DIALOG. <http://www.ferc.gov/>. Accessed 23 Sept 2013
17. Erol-Kantarci M, Kantarci B, Mouftah HT (2011) Reliable overlay topology design for the smart microgrid network. *IEEE Netw* 25(5):38–43
18. Wang D, Ren C, Sivasubramaniam A, Urgaonkar B, Fathy H (2013) Energy storage in datacenters: what, where, and how much? Available via DIALOG. <http://www.cse.psu.edu/bhuvan/papers/ps/sigmetrics12.pdf>. Accessed 23 Sept 2013
19. Goiri Í, Katsak W, Le K, Nguyen TD, Bianchini R (2013) Parasol and greenswitch: managing datacenters powered by renewable energy. In: *Proceedings of architectural support for programming languages and OS (ASPLOS)*
20. Cao Y, Jiang T, Zhang Q (2012) Reducing electricity cost of smart appliances via energy buffering framework in smart grid. *IEEE Trans Parallel Distrib Syst* 23(9):1572–1582
21. Urgaonkar R, Urgaonkar B, Neely MJ, Sivasubramaniam A (2011) Optimal power cost management using stored energy in data centers. In: *Proceedings of ACM special interest group on measurement and evaluation (SIGMETRICS)*
22. Yao Y, Huang L, Sharma A, Golubchik L, Neely M (2013) Power cost reduction in distributed data centers: a two time scale approach for delay tolerant workloads. *IEEE Trans Parallel Distrib Syst* 25(1):200–211
23. Neely MJ (2010) *Stochastic network optimization with application to communication and queueing systems*. Morgan & Claypool, San Rafael
24. Arno R, Friedl A, Gross P, Schuerger RJ (2012) Reliability of data centers by tier classification. *IEEE Trans Ind Appl* 48(2):777–783
25. Li J, Li Z, Ren K, Liu X, Su H (2011) Towards optimal electric demand management for internet data centers. *IEEE Trans Smart Grid* 2(4):1–9
26. Chen Y, Das A, Qin W, Sivasubramaniam A, Wang Q, Gautam N (2005) Managing server energy and operational costs in hosting centers. In: *Proceedings of ACM special interest group on measurement and evaluation (SIGMETRICS)*
27. Damousis IG, Alexiadis MC, Theocharis JB, Dokopoulos PS (2004) A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation. *IEEE Trans Energy Convers* 19(2):352–361
28. Carrión M, Philpott AB, Conejo AJ, Arroyo JM (2007) A stochastic programming approach to electric energy procurement for large consumers. *IEEE Trans Power Syst* 22(2):744–754
29. Palma-Behnke R, Benavides C, Lanas F et al (2013) A microgrid energy management system based on the rolling horizon strategy. *IEEE Trans Smart Grid* 4(2):996–1006
30. Lu L (2013) *Online energy generation scheduling for microgrids with intermittent energy sources and co-generation*. The Chinese University of Hong Kong, Hong Kong
31. San Diego's AIS Rides Out Power Outage. <http://www.datacenterknowledge.com/archives/2011/09/09/san-diegos-ais-rides-out-power-outage/>. Accessed 23 Sept 2013
32. Huang Y, Mao S, Nelms RM (2014) Adaptive electricity scheduling in microgrids. *IEEE Trans Smart Grid* 5(1):270–281
33. National Renewable Energy Laboratory (2012) *Variance analysis of wind and natural gas generation under different market structures: some observations*
34. Costa AM (2005) A survey on benders decomposition applied to fixed-charge network design problems. *Comput Oper Res* 32(6):1429–1450
35. Çakır O (2009) Benders decomposition applied to multi-commodity, multi-mode distribution planning. *Expert Syst Appl* 36:8212–8217
36. <http://generatorjoe.net/html/fueluse.asp>. Accessed 23 Sept 2013
37. Zhang Y, Gatsis N, Giannakis GB (2012) Robust management of distributed energy resources for microgrids with renewables. In: *Proceedings of IEEE international conference on smart grid communications (SmartGridComm)*

Energy Management of Internet Data Centers in Smart  
Grid

Jiang, T.; Yu, L.; Cao, Y.

2015, XIV, 102 p. 23 illus., Hardcover

ISBN: 978-3-662-45675-0