

Research Article

A Dynamic Thermal-Allocation Solution to the Complex Economic Benefit for a Data Center

Hui Liu ¹, Wenyu Song,¹ Tianqi Jin,² Zhiyong Wu,³ Fusheng Yan,¹ and Jie Song⁴

¹New Energy Science and Engineering Department, School of Metallurgy, Northeastern University, Shenyang 110819, Liaoning, China

²Thermal Engineering Department, School of Metallurgy, Northeastern University, Shenyang 110819, Liaoning, China

³China Railway Rolling Stock Corporation Zhuzhou Institute Co., Ltd., Zhuzhou 412001, Hunan, China

⁴Software Engineering Department, Software College, Northeastern University, Shenyang 110819, Liaoning, China

Correspondence should be addressed to Hui Liu; liuh@smm.neu.edu.cn

Received 17 March 2020; Revised 5 May 2020; Accepted 15 May 2020; Published 15 June 2020

Guest Editor: Baogui Xin

Copyright © 2020 Hui Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Data centers, which provide computing services and gain profits, are indispensable to every city in the information era. They offer computation and storage while consuming energy and generate thermal discharges. To maximize the economic benefit, the existing research studies on the data center workload management mostly leverage the dynamical power model, i.e., the power-aware workload allocation. Nevertheless, we argue that for the complex relationship between the economic benefit and so many attributes, such as computation, energy consumption, thermal distribution, cooling, and equipment life, the thermal distribution dominates the others. Thus, thermal-aware workload allocation is more efficient. From the perspective of economic benefits, we propose a mathematical model for thermal distribution of a data center and study which workload distribution could determinately change the thermal distribution in the dynamic data center runtime, so as to reduce the cost and improve the economic benefits under the guarantee of service provisioning. By solving the thermal environment evaluation indexes, RHI (Return Heat Index) and RTI (Return Temperature Index), as well as heat dissipation models, we define quantitative models for the economic analysis such as energy consumption model for the busy servers and cooling, energy price model, and the profit model of data centers. Numerical simulation results validate our propositions and show that the average temperature of the data center reaches the best values, and the local hot spots are avoided effectively in various situations. As a conclusion, our studies contribute to the thermal management of the dynamic data center runtime for better economic benefits.

1. Introduction

Data center (DC) is an information service platform with efficient equipment and perfect management mechanisms. Under the background of the high-speed information age, the global demand for Internet business and information services is increasing year by year. According to statistics, the global Internet Data Center (IDC) business market as a whole will exceed 130 billion dollars in 2020, and DCs around the world will consume 8% of the world's electricity and put a heavy environmental burden on society [1].

The huge energy consumption accompanies extremely low resource utilization, which is reported to be between 5% and 25% in typical DCs [2]. The main reason for this is that

DC operators often adopt redundant resource deployment strategies in pursuit of high performance, quality of service, and reliability. Therefore, all servers will be at the highest busy ratio regardless of the load. The low utilization of these multidimensional resources (CPU, storage, memory, and network bandwidth) leads directly to huge resource waste. At the same time, it also increases the cost of other supporting equipment such as cooling system and power distribution unit. Under this development trend, how to reduce the high cost and high energy consumption caused by the high demands has been concerned by scholars.

The key to the cost control of facility operation and maintenance in a DC is the electricity cost of the equipment,

and the main influencing factors are the running time, quantity, and distribution of high-power equipment [3]. In addition, as the main components of a DC, IT equipment and cooling equipment account for about 90% of the energy consumption of the DC, among which the cooling equipment accounts for about 40%–60% [4, 5]. In the optimization scheme, the optimized workload allocation is adopted to improve the cooling capacity and adjust the temperature distribution in the DC, so as to reduce the running time of the air-conditioning equipment and the electricity cost, reduce the hot spot of temperature, and further reduce the possibility of equipment damages and make the effect of full-life cycle cost control of the DC more obvious. Therefore, there is a very complex relationship between economic benefit and calculation, energy consumption, heat distribution, cooling, and equipment life, among which heat distribution dynamically dominates other properties [6].

It is necessary to analyze the economic effects of the improvements in energy efficiency achieved by existing methods. In this paper, the steady and transient numerical simulation is used to seek the switch strategy of energy-saving operation, and the two main problems of improving the airflow organization mode and busy servers distribution mode are studied. To improve the economic performance of the DC cooling strategy from the perspective of dynamic thermal environment, the following research contents are studied:

- (1) A reasonable calculation model and a calculation method are selected, the temperature and velocity distribution of airflow in the DC are simulated, and the temperature distribution and airflow velocity distribution of the specified section of the DC are obtained;
- (2) The influence of power distribution, workload allocation, and overall busy ratio on the thermal environment of the DC is studied, and the optimized switch strategy is analyzed;
- (3) The influence of power and cooling quantity changes on transient temperature rise of busy servers is analyzed, and transient switch strategy analysis is carried out under the condition of dynamic change of switch quantity.
- (4) Based on the heat distribution of the DC, the influence factors and the influence of thermal environment on the economic benefit of the DC are analyzed. The complex economic benefit problem, which is expressed as a nonlinear optimization problem with multiparameters and multiconstraints, is solved.

The rest of the paper is organized as follows. Section 2 introduces the related works about thermal environment and the economic benefit of DCs. Section 3 proposes the mathematical model and thermal evaluation metrics for modeling thermal environment. Section 4 gives workload allocation for servers according to the mathematical thermal model and evaluates the influence of busy ratio. Section 5 discusses the transient analysis of the dynamic switch strategy and Section 6

analyzes the economic effects. Finally, the conclusions and future works are summarized in Section 7.

2. Related Works

In this section, we provide a brief overview of some existing works on two major topics: the first is the thermal optimization for DCs, and the second is the economic benefit optimization for DCs.

The thermal environment of DCs has been studied in existing literature from the aspects of air distribution, construction of hot and cold enclosed aisles, and reduction of inefficiencies of the equipment. In terms of air distribution, many researchers have conducted a lot of studies on air supply methods, perforated tiles, equipment layout, and so on. Chu et al. [7] studied the thermal influence of air intake flow and inlet layout on the DC with the enclosed cold aisle. They found that the uniformity of inlet flow did not improve with the strength of flow, the power consumption increased instead. However, when the inlet direction was deflected to the rack direction, the flow uniformity in the rack was greatly improved. In terms of cold aisles design, Cho and Woo [8] designed the new row-based cooling system, in which the air conditionings, racks, and cold and hot aisles are all enclosed in one line. The cooling efficiency of this model, RHI, and RTI are 20%, 73.2%, and 50% higher than those of the open aisle, respectively. In terms of reducing the energy consumption of equipment, Jin et al. [9] summarized the important influence of the accurate server's power model on the energy saving and reliability of the DCs. In the model, the effects of cooling output, inlet temperature, and energy saving of the server are considered. It is pointed out that the peak power of the server accounts for 40%–50% of the rated power. Energy-saving technology can reduce idle power from 55% of rated power to 15%. With the improvement of heat dissipation requirements, the placement of racks in DC has gradually developed from the extensive form to the orderly form, that is, the racks are arranged in order to naturally form cold aisles and hot aisles [10, 11]. Because the enclosed cold aisles can effectively contain the cold flow loss caused by mixing of cold and hot air in the cold aisles, it has been gradually accepted and popularized in engineering practice [12].

In most studies, the modeling studies focused strictly on steady-state analysis of DCs, and fixed workload allocation and rack powers are imposed. For transient analyzes, the switch coefficient of servers is crucial as it significantly affects the amount of time it takes to reach the steady state [13]. Therefore, in terms of the energy efficiency optimization, we are interested in the thermal impact of various switch strategies [14, 15].

The maximization of economic benefit for DCs is a complex problem because it is with many factors and restrictions. Classic solution only takes one-time investment and operation cost into consideration [16]. The former includes investment on servers, network, accessories, buildings, power generators, and computer room air conditionings (CRAC). Such investment is static and is not

considered in our researches. The latter mainly includes cost of equipment's maintenance and energy. Such cost is dynamic and is considered in our researches. The sophisticated economic benefit models can effectively estimate and even predict the total investment and annual operation cost of a DC. Besides, economic benefit optimization can be reached by several approaches, for example, a model based on a resource management technique and semi-Markov decision process [17], green scheduling for cloud data centers in an economical way by renewable energy trading with the power grid [18], scheduling of data-oriented tasks in geographically distributed cloud data centers [19], transmission cost reduction from the perspective of DC users [20], and replica factor dynamical adjustment to reduce the resource consumption and guarantee the economic profit [21]. However, to the best of our knowledge, there are few researches leveraged the thermal management for economic benefit optimization. There are few researches modeled economic benefit as a maximization problem and took the thermal distribution, profits, energy cost of servers, and cooling system into consideration. Otherwise, the economic benefit models are very challenging to be accurate, quantitative, and efficient. The thermal optimization is mainly associated with energy optimization, for example, MirhoseiniNejad et al. [22] considered thermal effects of server workloads, which in conjunction with control parameters of the cooling unit, save more power than optimizing each of them separately.

3. Mathematical Model and Thermal Evaluation Metrics

In this paper, the airflow distribution model of a DC is "airflow supplied from the lower side and returned from the upper side via the enclosed cold aisle", as shown in Figure 1. Comparing with other airflow modes, our choice has obvious advantages [23, 24]. In Figure 1, all dimensions are listed in $X \times Y \times Z$ order. The size of the DC is $14 \times 12 \times 2.5$ (m), in which there are 4 (columns) \times 11 (racks) with the size of $0.6 \times 1.1 \times 2$ (m). The size of the enclosed cold aisle is 6.6×1.2 (m) and width on each side of hot aisles is 1.4 (m). The size of the server is $0.5 \times 1 \times 0.0495$ (m) and the CRAC is $0.9 \times 1.92 \times 2$ (m). The size of the front rack is $1.82 \times 1.1 \times 2$ (m). In addition, the cooling inlet and outlet of the rack are fully open, and they are in the same size of 0.6×2 (m), and the size of each air conditioner's outlet is 0.8×0.9 (m). The air at 17°C supplied from the floor inlet refrigerates servers through two columns of standard 42 U racks and then returns to the CRAC on top of the hot aisle; the cold aisle is enclosed. There are two air outlets placed on top of each air conditioner. The ambient temperature is set at 20°C . The average temperature difference of return air is predicted to be 10°C . When air flows through the rack, the cooling loss can be considered as sensible heat exchange, which could be calculated by equation (1) to predict the air supply volume in numerical simulation:

$$Q = c_p G \Delta T, \quad (1)$$

where c_p is the specific heat capacity of the fluid at constant pressure (J/kg·K), G is the air volume (kg/s), and ΔT is the average temperature difference of return air ($^\circ\text{C}$).

We design an extreme condition that 20 highly integrated servers (1 U for each) are distributed in each rack in the model. The cooling capacity in this paper depends on the heat conversion rate of IT equipment, which is about 80%. In the numerical simulation model, the rated power of the 1 U server is 200 W/U and the idle power is 20 W/U. Figure 2 shows the heat load of the DC with different busy ratios and the designed cooling flow rate.

In this paper, thermal analysis software IcePak is used for numerical simulation analysis. The finite element volume method is used for the discrete form and the SIMPLE algorithm is used for solution. The minimum cell's size was $3 \text{ mm} \times 1 \text{ mm} \times 6 \text{ mm}$, and the total number of grids was 1.9×10^6 . As for the selection of turbulence models, the zero equation has the advantage of less computation, obtaining more accurate results than other models [25]. The assumed conditions are as the follows:

- (1) The low-speed air in the DC can be regarded as Newtonian fluid, and the dissipation work caused by the viscous force of the fluid can be ignored
- (2) The fluid domain is steady turbulence
- (3) The nonslip boundary condition is applied on the air inside the DC
- (4) The air tightness of the DC is good, and the influence of air leakage is ignored
- (5) According to the Boussinesq hypothesis, the change of fluid density only affects the buoyancy

In this paper, a three-dimensional incompressible fluid model is established, which follows the laws of mass conservation, momentum conservation, and energy conservation. Combined with the Reynolds-averaged Navier-Stokes method, the governing equations are as follows [26]:

- (1) Continuity equation

$$\frac{\partial \rho}{\partial t} + \rho \frac{\partial \bar{v}_i}{\partial x_i} = 0, \quad (2)$$

where t is the time (s), \bar{v}_i , \bar{v}_j are vector speeds (m/s), and x_i , x_j are vector coordinates (m).

- (2) Momentum equation

$$\rho \left(\frac{\partial \bar{v}_i}{\partial t} + \bar{v}_j \frac{\partial \bar{v}_i}{\partial x_j} \right) = \rho f_i - \frac{\partial \bar{p}}{\partial x_i} + \frac{\partial}{\partial x_j} \left(\mu \frac{\partial \bar{v}_j}{\partial x_j} - \rho \overline{v'_i v'_j} \right), \quad (3)$$

where f_i is the body force (N), μ is the dynamic viscosity of the fluid (Pa·s), and ρ is the fluid density (kg/m^3). It is worth noting that $\rho \overline{v'_i v'_j}$ in momentum equation (3) is the turbulent stress. In order to determine its value, it needs to be solved by combining the turbulent zero-equation mode, i.e., equation (4).

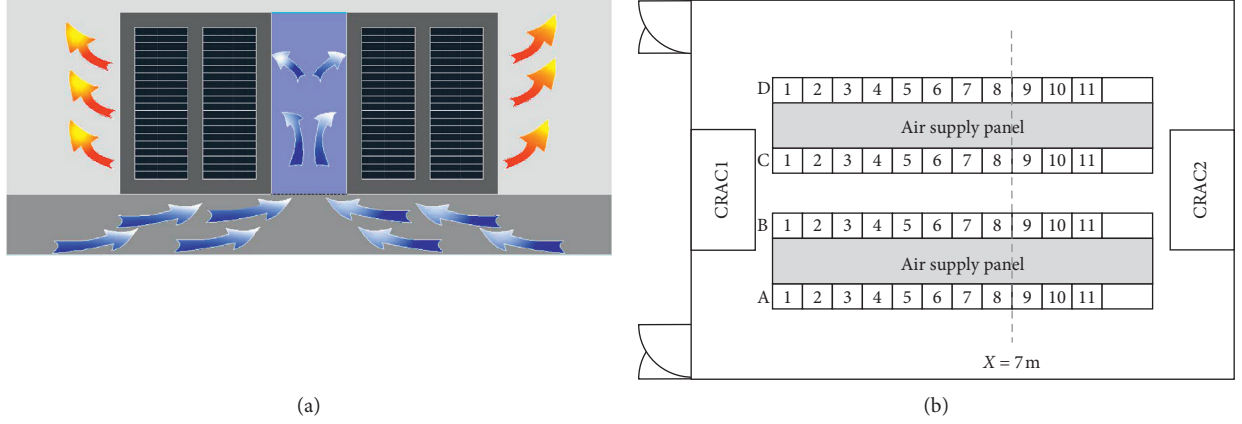


FIGURE 1: Physical model of the DC. (a) Air supply mode. (b) Diagram of the physical mode.

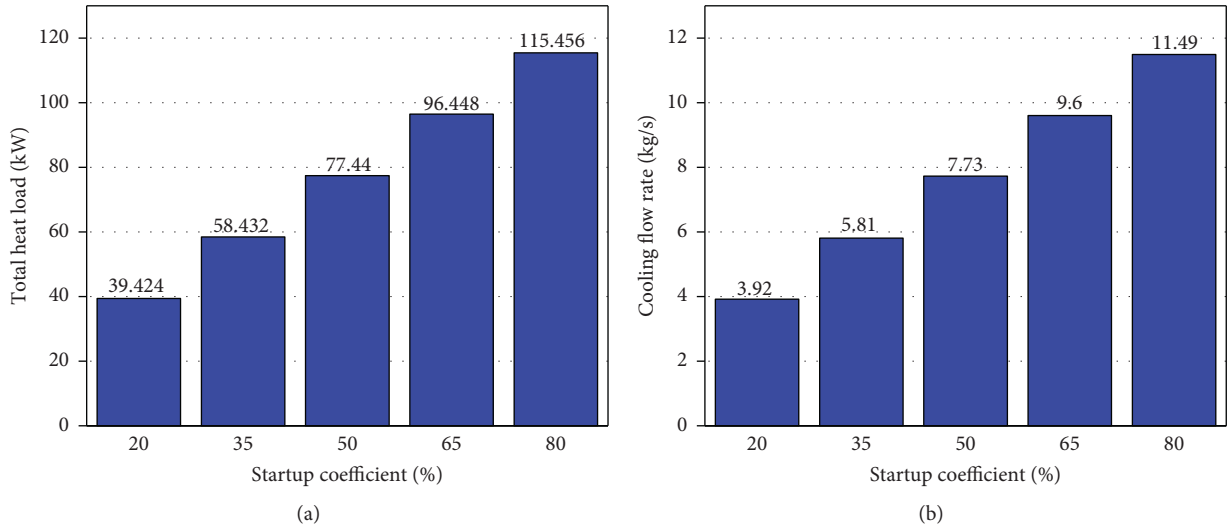


FIGURE 2: Total heat load (a) and cooling flow rate (b) of the DC.

The turbulence model: zero-equation mode

$$\begin{aligned} \tau_t &= -\overline{\rho v_i' v_j'} = v_T \left(\frac{\partial \bar{v}_i}{\partial x_j} + \frac{\partial \bar{v}_j}{\partial x_i} \right) \\ &= \rho l^2 \sqrt{\frac{1}{2} \left(\frac{\partial \bar{v}_i}{\partial x_j} + \frac{\partial \bar{v}_j}{\partial x_i} \right)^2 \left(\frac{\partial \bar{v}_i}{\partial x_j} + \frac{\partial \bar{v}_j}{\partial x_i} \right)}. \end{aligned} \quad (4)$$

(3) Energy conservation equation

$$\frac{\partial(\rho \bar{T})}{\partial t} + \frac{\partial(\rho \bar{v}_i \bar{T})}{\partial x_i} = \frac{\lambda}{c_p} \frac{\partial^2 \bar{T}}{(\partial x_j)^2} - \overline{\rho v_i' T'} + S, \quad (5)$$

where T is the fluid temperature ($^{\circ}\text{C}$), λ is the thermal conductivity of the fluid ($\text{W/m}\cdot\text{K}$), S is the source item, and $\overline{\rho v_i' T'}$ is the turbulent thermal diffusion term, the expression of which is shown in the following equation:

$$\overline{\rho v_i' T'} = \frac{v_T}{\text{Pr}_t}, \quad (6)$$

where v_T is the turbulent viscosity ($\text{Pa}\cdot\text{s}$), Pr_t is an empirical coefficient, the value of which is greatly affected by the material thermal properties and turbulent intensity. l is the mixing length (m).

Equations (7)–(11) give the boundary conditions of the model:

$$v_{z=0} = v_0, \quad (7)$$

$$p_{z=H} = p_0, \quad (8)$$

$$T_{air}|_{z=0} = T_0, \quad (9)$$

$$q_w = 0, \quad (10)$$

$$\lambda \frac{\partial T_{air}}{\partial x_i} \Big|_{x_i=x_1} = q = \text{const} \quad (11)$$

Combined with the governing equations and boundary conditions in this model, it can be seen that the velocity vector in the energy equation can be obtained by solving continuity equation (2) and momentum equation (3) of a certain position under the initial conditions. Furthermore, the turbulent heat diffusion value in the energy equation is obtained by the turbulent stress solved under the specific condition and the given turbulent Prandtl number, and on this basis, we can obtain the value of turbulent thermal diffusion and temperature at a certain point in this system. This process is iterated to gradually generate the velocity field and temperature field of the whole space.

A verification model was established according to the experiment of Arghode et al. [27], consistent with experimental conditions in this paper. The cold aisle temperature measured by the temperature monitoring car of racks 2–7 and racks 9–12 was compared with the simulated results, as shown in Figure 3.

Figure 3 shows experimental and simulated temperature of the cold aisle. The error of racks in the middle position is the minimum, and the mean absolute error is about 5%. It can be seen that the error is caused by uncontrollable factors between the assumed condition and the actual situation. The reason for the higher temperature of the lower servers may be that the temperature measuring car blocks the air flow in the cold aisle. The temperature shows mostly good agreement with the numerical results. It can be considered that the model is feasible.

In order to evaluate the selected scheme, RHI (Return Heat Index) and RTI (Return Temperature Index) are introduced to conduct comprehensive evaluation [15, 23], where RHI represents the utilization ratio of the air-cooling capacity of the rack, and RTI evaluates the air distribution in the DC. Details are shown in Table 1.

RHI can be used to evaluate the thermal environment and determine the existence and specific location of local hot spots, while RTI can be used to objectively evaluate the airflow distribution of the rack and the overall DC.

4. Influence of Workload Allocation on Complexity Thermal Environment

The influence of workload allocation and distribution of racks on the heat dissipation of the DC and the influence of parameters of the DC change on the optimization of heat dissipation under nonfull workload conditions are studied in this section.

4.1. Workload Allocation for Servers. According to the current common mode of busy servers allocation in DCs, 8 classic allocation modes of busy servers were designed and are shown in Table 2 and Figure 4, named by the workload allocation of the servers, and Figure 4 shows that there are

significant differences of thermal environment in the DC. There are two reasons for this phenomenon. First of all, when the cold air flows through the enclosed aisle, there is different cooling attenuation at different workload allocation of servers. In addition, under different workload distribution modes, the reexchanging heat amount of airflow with different cooling attenuation in the hot aisle is also different. The cold energy utilization in the centralized distribution models is significantly lower than that in the decentralized model. The hot zone exists in the DC with the centralized distribution model, as shown in Model 1~6, the uniformity of thermal environment of Model 7 and Model 8 is better, and the temperature difference is about 2°C. The results show that whether the thermal environment is good or not depends on the mixing degree of cold airflow in the open hot aisle, the better the heat transfer performance, the more uniform the temperature distribution and the better the overall heat dissipation. Secondly, because the trend of the vertically upward air supply in the enclosed aisle is greater than that of the side direction rack fan, less cold air flows through the lower part of the rack's servers than the upper part. This conclusion can be confirmed by the temperature comparison of Model 1, Model 2, and Model 3 in Figure 5, and heat transfer capacity is weakened successively from Model 2 and Model 3 to Model 1. In addition, the hot zone appeared in models where the workload was distributed near the lower layers, such as Model 1, Model 4, and Model 6.

According to GB50174-2017 of China, the standard range of safe operation in DCs is 18–28°C. When the busy ratio of the DC is 0.5, the temperature comparison of the overall thermal environment in 8 cases is shown in Figures 5(a)–5(h). The peak temperatures of Model 1/4/6/7 exceed the upper limit of 28°C; it will inevitably increase the consumption of refrigerating capacity and electricity cost for the stable operation. The peak temperatures of Model 2/3/5/8 are within the standard range. It indicates that an economical operation mode for DCs is to arrange the busy servers close to the middle and upper layers of the rack in the process of nonfull load operation. In addition, busy servers in Model 5, Model 2, and Model 3 are all distributed in the middle and upper layers, and they are relatively more dispersed in Model 5. Comparing the thermal environment of Model 2/3/5/8, the thermal stratification of Model 5 is more uniform, and the average temperature of the thermal environment is the lowest. On the basis of centralized distribution, proper dispersion is beneficial to improve the cooling effect, but it is not that the more dispersed the better. For example, the thermal environment of Model 8 is not as good as that of Model 5, because the servers' excessive dispersion will increase the possibility that the servers are distributed in the lower rack, and it is not conducive to cooling.

According to Table 1, RHI and RTI are obtained to evaluate the local airflow structure comprehensively. Figure 6(b) shows that RHI is relatively high in the mode that the busy servers are far away from the lower layers, for example, in Model 2, 3, and 5. It means that in nonfull load conditions, the middle and upper layout modes of the busy servers can effectively reduce the mixing of hot and cold air

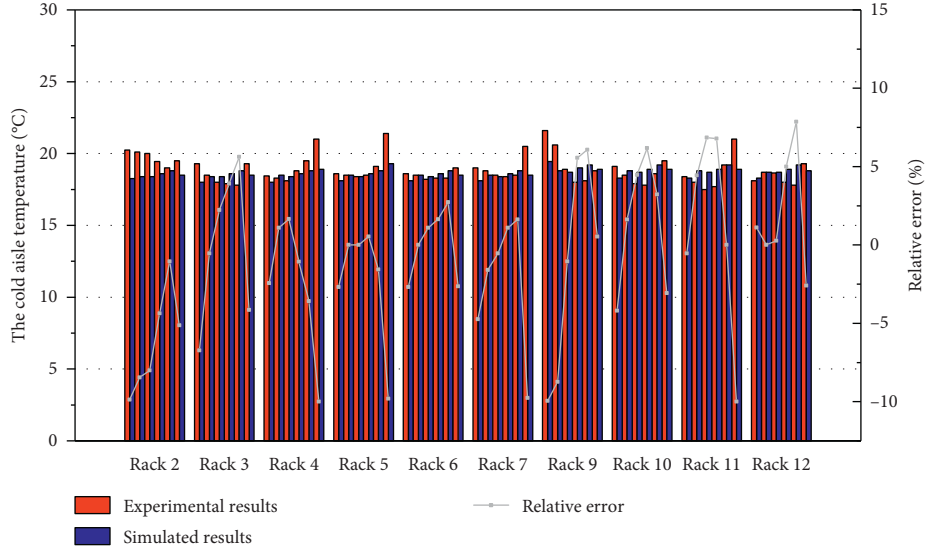


FIGURE 3: Temperature comparison of experimental and simulation results in cold aisles.

TABLE 1: Description of two thermal evaluation indexes.

Thermal evaluation index	RHI	RTI
Expression	$Q/(Q + \delta Q)$	$(T_{out} - T_{in}/\Delta T) \times 100\%$
Evaluation criterion	RHI $\in (0, 1)$. The more the RHI approaches 1, the better the heat transfer, the less the mixing of cold and hot air, and the higher the utilization rate of air-cooling capacity is.	The more the RTI approaches 1, the better the air distribution is. When RTI > 1 , there is hot air backflow. While RTI < 1 , it indicates cold air bypass.

TABLE 2: Allocation modes of busy servers.

Workload allocation model	Busy servers distribution
Model 1	Concentrate in the lower layers
Model 2	Concentrate in the middle layers
Model 3	Concentrate in the upper layers
Model 4	Central free
Model 5	Upper alternating segment
Model 6	Lower alternating segment
Model 7	Alternating segment
Model 8	Single alternating

and improve the utilization of cooling capacity. Through the relative temperature index RTI, we can find that RTI of Model 7 is closest to 1, followed by Model 5. However, based on the abovementioned conclusions, in the three models with better RHI of Model 2, 3 and 5, RTI of Model 5 is closest to 1. The maximum temperature of Model 5 is about 7°C lower than Model 7, that is, Model 5 is the best model for the comprehensive evaluation of the heat and air distribution.

4.2. Influence of Busy Ratios. Based on the abovementioned conclusions, Model 2, 3, 5, and 8 are adopted to analyze the variation regularity of thermal environment performance indexes RHI and RTI, and the switch strategy on energy saving will be optimized. When the busy ratios are, respectively, 0.2, 0.35, 0.5, 0.65, and 0.8, the numbers of busy servers are, respectively, 4, 7, 10, 13, and 16. Figure 7 shows

that RTI increases with the increase of the busy ratio, and the range gradually increases from about 0.75 to 1.3. This can be attributed to that when the busy ratio is small, the subcooled zones are large, and the loss of bypass cold air occurs. As the busy ratio increases, the hot zones gradually appear instead of subcooled zones, the possibility of hot air recirculation is increased. Secondly, RHI decreases with the increase of the busy ratio. This is because the increase of the busy ratio promotes the possibility of hot air backflow, which leads to hot air mixing with cold air in the enclosed cold aisle, reducing the cooling efficiency of air supply. It can be seen that given the ideal cooling amount making the thermal environment at the same temperature gradient, RHI and RTI are closest to the ideal value at the busy ratio of 0.5, and the cooling efficiency is the highest. When the busy ratio is more than or equals to 0.5, RTI of Model 5 is closest to ideal value 1, RHI is the largest relative to other models, and the overall thermal environment is optimal. When the busy ratio is less than 0.5, Model 2 is the best choice.

5. Transient Analysis of Dynamic Switch Strategy

There are numerous transient scenarios in the actual operation of DCs. Therefore, the dynamic switch of workload is very important for the stable operation of DCs. In addition to the influence of the dynamic switch degree of the workload quantity on the cooling performance of servers,

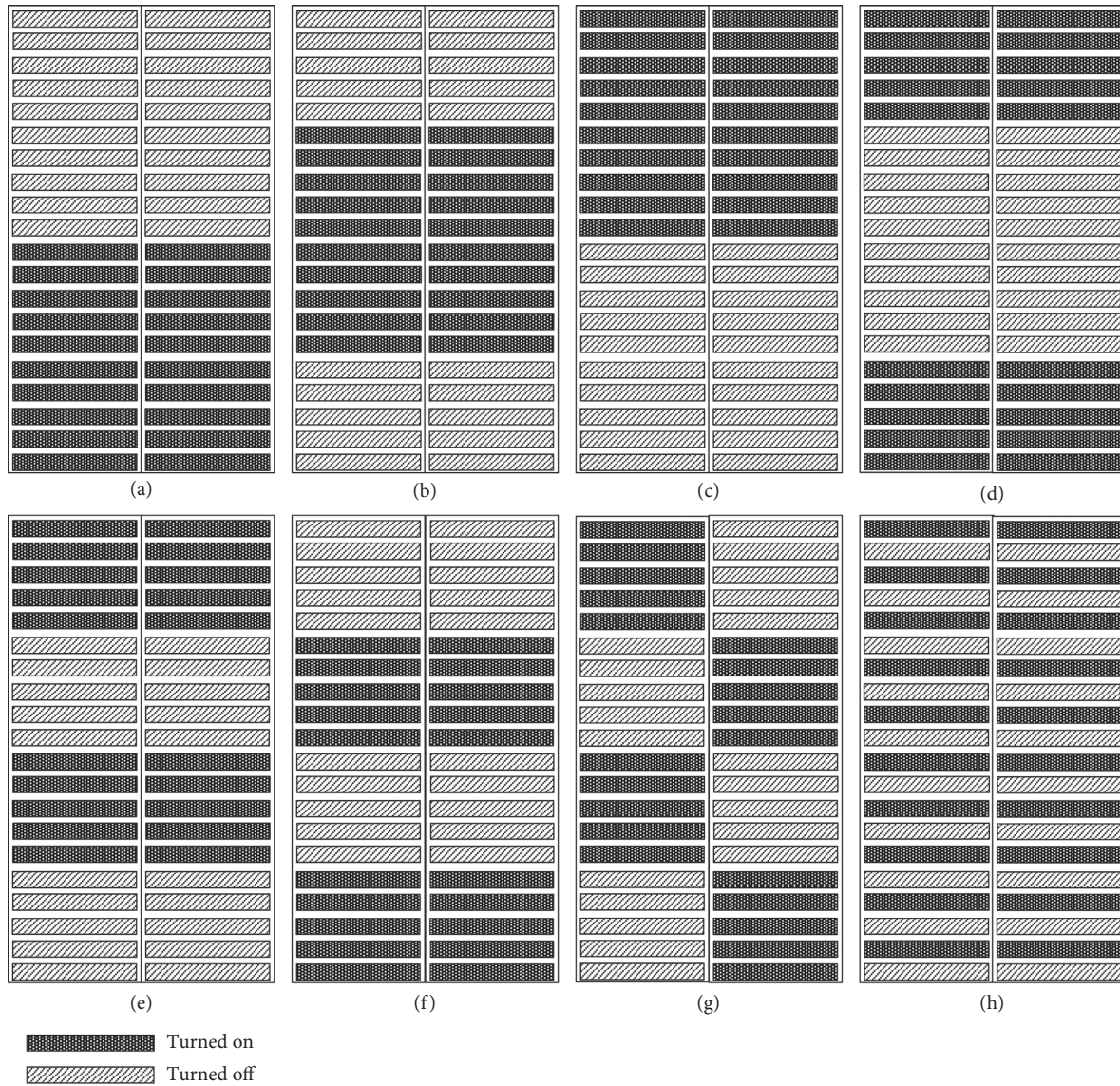


FIGURE 4: Schematic diagram of the workload allocation scheme. (a) Model 1. (b) Model 2. (c) Model 3. (d) Model 4. (e) Model 5. (f) Model 6. (g) Model 7. (h) Model 8.

the location change of busy/idle servers will also affect the transient thermal environment of DCs. In transient modeling analysis, the switch sequence of busy servers in different positions will not only affect the generation location of time-varying heat but also cause differences in cold and hot air mixing at the outlet.

Four classical cases were set up with different quantities of busy servers and different dynamic switch sequences for numerical simulation to analyze their effects on the thermal environment of DCs. In Case C_0 , the total task load is evenly distributed on each server of the rack. In other three cases, the standard 42 U rack shall be divided equally into the upper, middle, and lower layers. By adjusting the dynamic switch sequence of the servers in the upper, middle, and lower layers, the influence of busy/idle states changing of servers on the transient thermal environment will be analyzed. The busy servers in Case C_1 are started in the order of

lower-upper-middle layers as the task load increases, and each server keeps a state of full load of 200 W. Busy servers in Case C_2 are started in the order of middle-upper-lower layers, and in Case C_3 they are started in the order of upper-middle-lower layers. The total power is evenly distributed to each server in proportion, and the transient curve of power changing with time is shown in Figure 8. The total power of a single rack increases from 1.2 kW to 2.6 kW and then to the full load of 4 kW, there are two instantaneous uprush of the power at 300 s and 600 s, and the corresponding cooling capacity is increased from 30% to 65% to 100% of cooling capacity requiring at full load, and according to the simulation results, the curve of the average temperature at the 9th rack that exits in column A at the central section of the DC over time is shown in Figure 9.

From the comparison of the four groups of data, it is concluded that sudden changes in power cause temperature

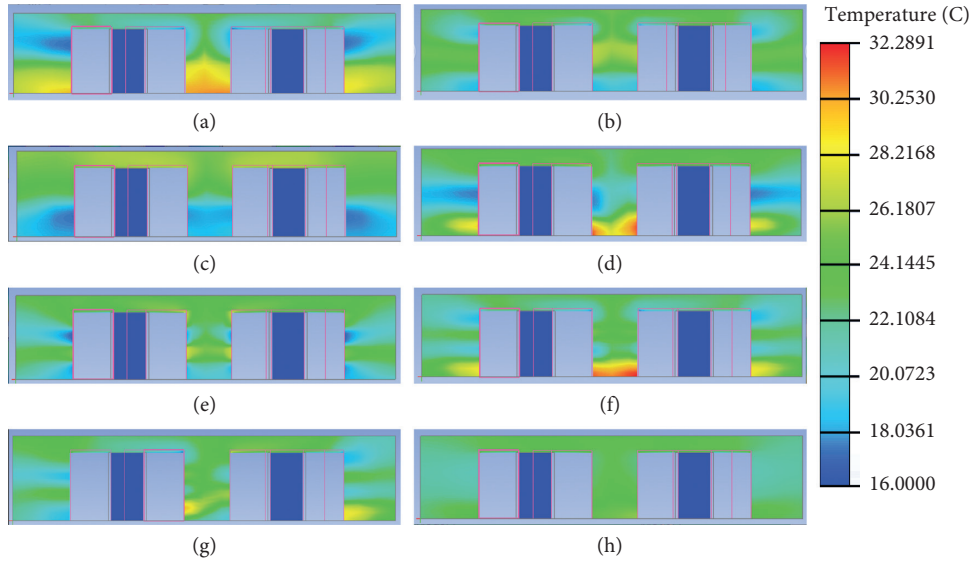


FIGURE 5: Temperature cloud diagram of different workload allocation at $X = 7$ m. (a) Model 1. (b) Model 2. (c) Model 3. (d) Model 4. (e) Model 5. (f) Model 6. (g) Model 7. (h) Model 8.

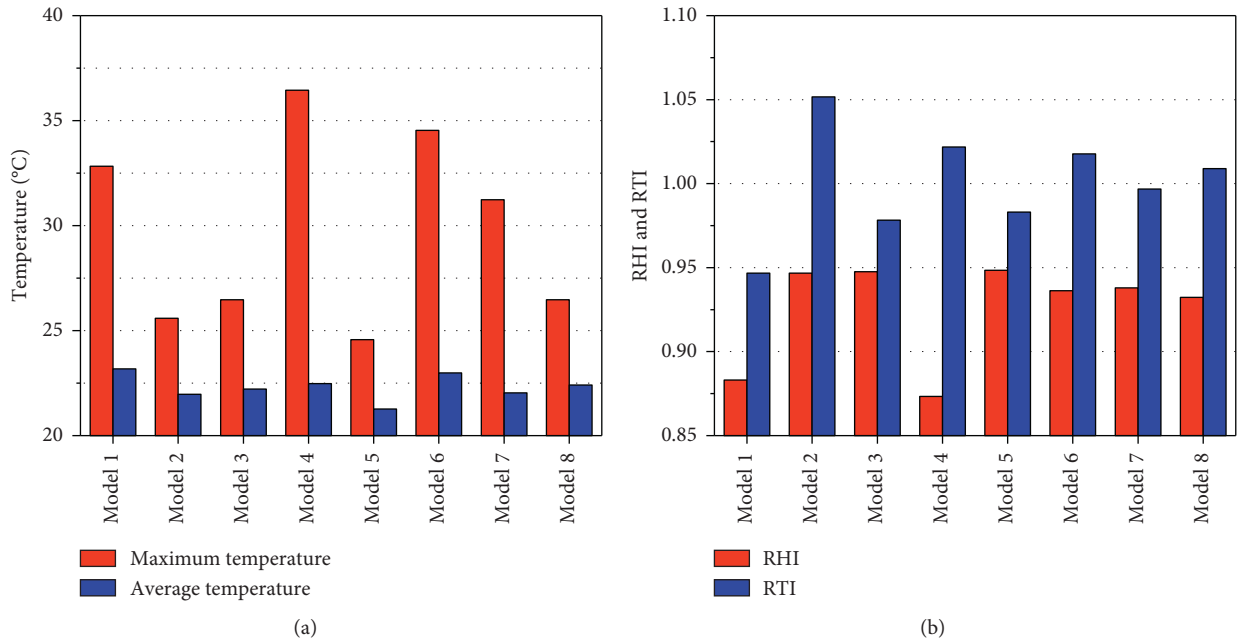


FIGURE 6: Evaluation based on temperature, RHI, and RTI of eight cases. (a) Temperature of thermal environment. (b) RHI and RTI.

fluctuations and result in the different transient effects in four cases. Considering the temperature fluctuation caused by different transient scenarios, when the load task increases, Optimal distribution of average temperature and the minimum temperature overshoot fluctuation occur in Case C_3 . While the worst temperature distribution and the strongest temperature overshoot occur in Case C_2 . The average temperature of the rack outlet with servers operating at full load finally remains the same in four cases, but the time for the average temperature to reach stability is different. This

indicates that, firstly, the dynamic change of the busy servers' location will lead to the circulation and mixing of hot and cold air around the rack outlet, and the cooling performance of different busy server locations will also be different. Secondly, as long as the parameters of the final state are consistent, changing the starting position and sequence of the servers will affect the time to reach the final stability without affecting the final thermal environment temperature. When the servers' workload increases, Case C_3 is undoubtedly a safe and feasible ideal switch strategy.

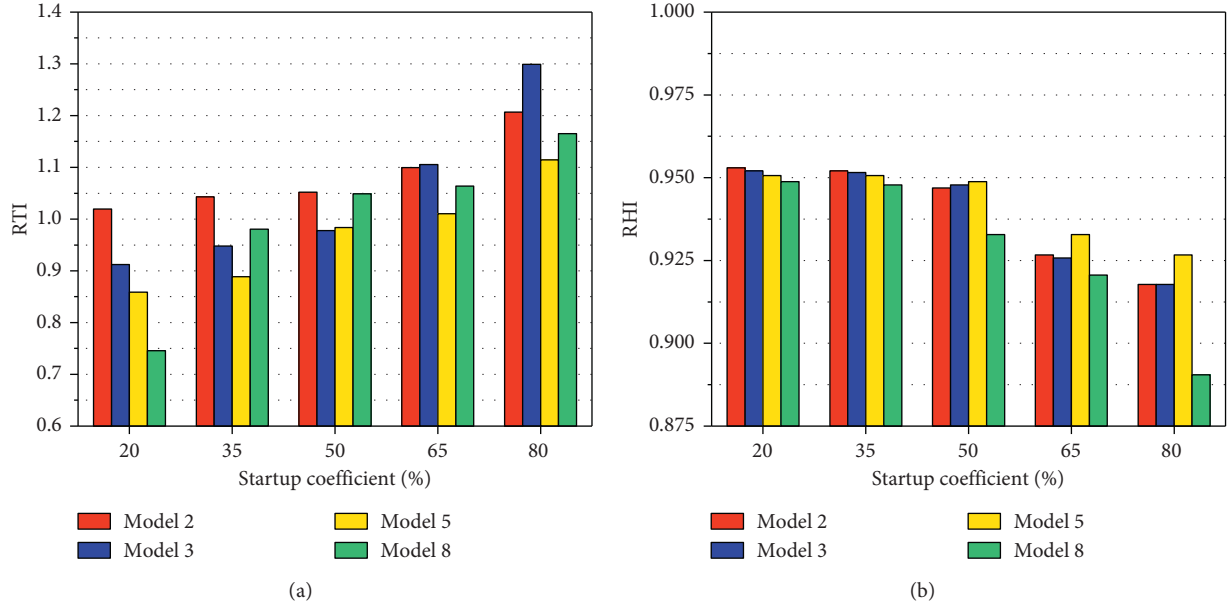


FIGURE 7: (a) RTI and (b) RHI in the DC with different busy ratios.

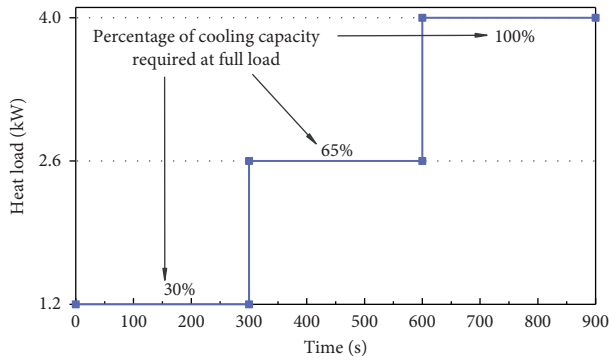


FIGURE 8: Schematic diagram of the transient attribute input file for changing the boot position.

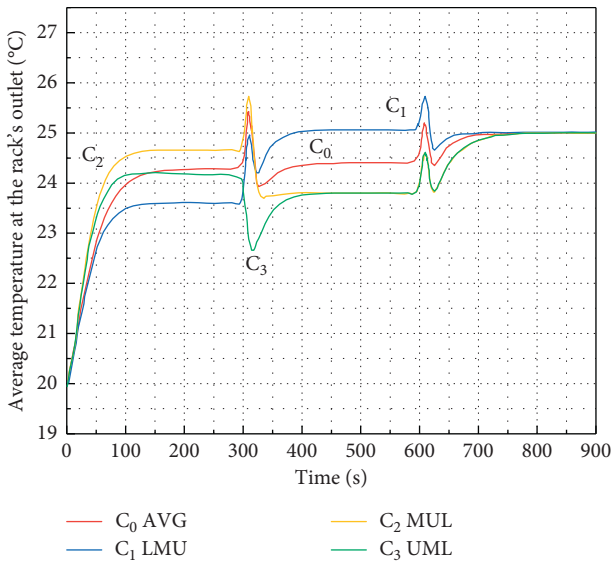


FIGURE 9: The average temperature at the outlet of the rack over time.

6. Economic Analyses

The economic benefits of a DC mainly include income and operating costs. DCs derive their revenues from the services they provide, and their costs are mainly energy bills. Energy costs include two parts: energy consumption for servers and cooling systems. Both of them are relevant to the number of servers. Meanwhile, the number of servers determines the request execution status, service quality, and server distribution mode of the DC. The former indirectly affects the benefits of the DC, while the latter indirectly affects the heat distribution and cooling costs. The relationships between the economic benefits and the relevant attributes are shown in Figure 10 including the symbols used in the section. We model and analyze the economic benefits of the DC with the following steps: firstly, the server's number of the DC is determined; secondly, the energy consumption of servers is determined, and then the energy consumption of the DC refrigeration is determined by the methods mentioned in previous sections; finally, the energy price and benefits of the DC are modeled on this basis, and the DC economic benefits are determined.

6.1. Energy Consumption for Busy Servers. Let a DC need n servers in the scheduling time t , and the time for the servers to process the request is composed of two parts: average wait time and average processing time for requests. θ is the average processing speed for the request (request/sec). γ is the arrival speed of the requests in scheduling time t (request/sec). p is the full load power of a server (watts). δ is the average waiting time for requests in t time slot (sec). R_0 is the maximum delay constraint of time of the requests (sec). q is the quality of service, i.e., the probability of task being immediately dealt with; the higher the probability, the higher quality of the service. Under the constraints, the average

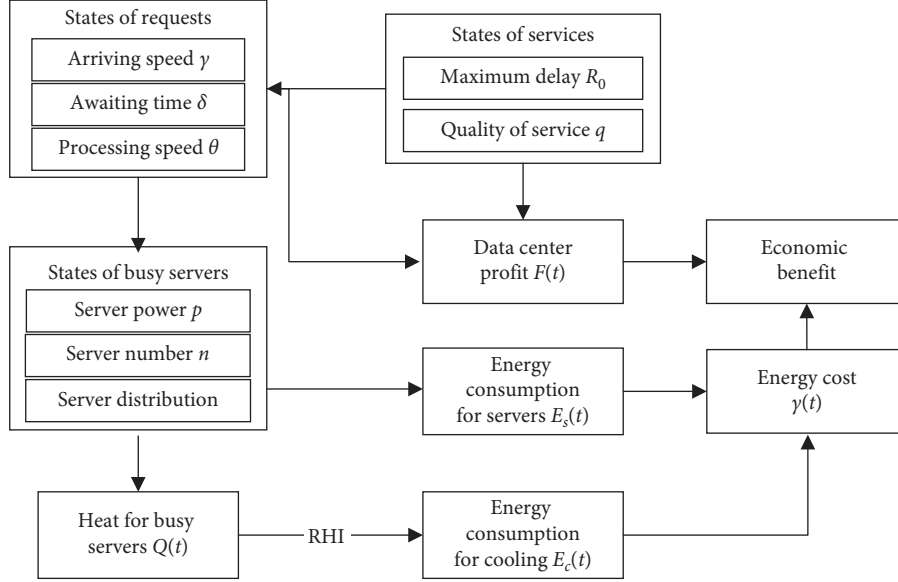


FIGURE 10: Relationships of benefits and relevant attributes.

waiting time for request is $\delta = q/(n \cdot \theta - \gamma)$, and the average response time for request R is $(1/\theta) + \delta$. Requiring $R \leq R_0$, so the following inequality is defined:

$$\frac{1}{\theta} + \frac{q}{n \cdot \theta - \gamma} \leq R_0. \quad (12)$$

Therefore, the constraint on the number of servers in the DC and the energy consumption under this constraint are as the follows:

$$E_s(t) = p \cdot n \cdot t, \quad \text{where } n \geq \frac{q}{R_0 \theta - 1} + \frac{\gamma}{\theta}. \quad (13)$$

6.2. Energy Consumption for Cooling. Based on the energy conservation law and the proposed optimization on cooling efficiency RHI, we study the energy consumption for cooling. In condition of n busy servers at time slot t , let $E_c(t)$ be the energy consumption for cooling, and $Q(t)$ be the heat produced by n servers, then $E_c(t) = Q(t) \text{RHI}^{-1}$, in which RHI is relevant to the switch strategy. When the number of running servers in the DC is n , the optimal switch strategy can be determined according to the heat distribution models in previous sections. For example, when n is half of the number of servers in the DC, the overall thermal environment of Model 5 in Table 2 is the optimal.

The mathematical expressions of $Q(t)$ can be defined using the lumped RC thermal model [28]. In the RC (resistor, capacitor circuit) model, the inside and outside of the server is considered as a heat transfer system with a certain temperature difference and thermal resistance. T_0 is the internal temperature of the servers, T_{amb} is the external temperature of servers, P is the server power, R is thermal resistance, and C is the heat capacity of the server. After t time the server temperature is shown in equation (14), and $Q(t)$ is shown in equation (15), where the parameters of equation (15) refer to the parameters in equation (12). The

refrigeration energy consumption $E_c(t)$ in the DC can be obtained.

$$T = RC(t) = PR + T_{\text{amb}} + (T_0 - PR - T_{\text{amb}}) \cdot e^{-t/RC}, \quad (14)$$

$$Q(t) = c_p G(T - T_0). \quad (15)$$

6.3. Energy Price. We utilize two different energy pricing models for energy deficient situation and energy adequate situation. Because more and more DCs apply the renewable energy, $Y(t)$ represents the time-dependent energy price for grid energy and the grid operator. We consider an exponential model [29] in energy deficient situation and flat-rate price for energy adequate situation, as shown in the following equation:

$$Y(t) = \begin{cases} y_0 e^{-\omega(t)}, & \text{if } t > t_\omega, \\ y_0, & \text{otherwise,} \end{cases} \quad (16)$$

where $\omega(t)$ is a normalized positive value of time t and t_ω represents the time slot when $\omega(t) = 0$, i.e., energy adequate situation, hence the energy price is constant y_0 .

6.4. DC Profit. The DC provides computing services for multiple users. These users share the same infrastructure, e.g., a user can share a DBMS with another in the context of databases. In return, each user pays the rent for resources to the provider according to the ‘‘pay as you go’’ model, i.e., a user only pays what it consumes [21]. Therefore, the DC’s profit in time slot t is relevant with w , R_0 , and q which are previously explained. Here, we use a logarithmic utility model, which follows the law of diminishing marginal utility and is widely used in a previous work [30]. $F(t)$ represents the time-dependent profit when the DC provides the services

with given average waiting time w , maximum delay time R_0 , and the probability of a request been processed q . $F(t)$ is calculated as the following equation:

$$F(t) = \varepsilon_0 \cdot \left(1 + k \cdot \log \frac{q}{R_0}\right) \cdot e^{-\text{ReLu}(w+\mu-R_0)}, \quad (17)$$

where ε_0 is the pricing constant, k is the adjustment coefficient, $\text{ReLu}()$ is the activation function, $\log q/R_0$ represents the contract profit which increases with expected service quality, and $e^{-\text{ReLu}(w+\mu-R_0)}$ represents the real service quality which decreases with the response delay.

6.5. Problem Definition. Aforementioned models define the problem of maximized economic benefits with the constraints. In a time slot t , the economic benefits are drawn from equations (13)–(17) as $F(t) - Y(t)[E_s(t) + E_c(t)]$, in which only n and RHI are variables, and the rest of the parameters are all constants. Besides, n and RHI are also relevant. Our heat distribution study ensures an optimal RHI for an n value. Therefore, we finally formalize the complex problem of maximizing economic benefit of the DC to the easily solved problem of the maximum value of a simple function. For long-term duration, the economic benefit is the aggregated value of each time slot t . In such conditions, each time slot in the duration is the schedule interval.

In conclusion, with the proposed model for the thermal distribution of a DC and the server distribution which could determinately change the thermal distribution, we define quantitative models for economic benefit and the relevant attributes such as energy consumption of servers and cooling system, energy price, profit, and service qualities. For maximizing the economic benefits, we remain the number of busy servers and thermal evaluation index as the schedulable attributes.

7. Conclusions

Considering the same air condition, the thermal steady-state numerical simulation is used to study the switch strategy of busy servers, and the thermal transient numerical simulation is used to study the switch strategy under transient changes in power of DCs. The conclusions are as follows:

- (1) When DC is running stably, there is a thermal stratification phenomenon in the cooling process. The thermal environment performance index is used for evaluation, and the decentralized distribution of the servers is conducive to heat dissipation. For the stable running conditions, when the busy ratio is at 0.5 or above, the upper alternating segment model has the best heat dissipation performance. For the lower busy ratio of less than 0.5, the mode that the servers concentrate in the middle layers is the best choice. Under the same temperature gradient, the cooling efficiency is the highest when the busy ratio is about 0.5.
- (2) The transient temperature characteristic under the dynamic switch of load condition is studied through

transient numerical simulation, and the effect of power change and cold quantity change on the transient characteristics of temperature is analyzed. The results show that the mode of servers in racks starting in the order of upper-middle-lower layer is an ideal strategy to meet the safety and feasible operation of DCs with enclosed cold aisles.

- (3) The influence of thermal environment on the economic benefits of DC is analyzed. We quantified the relationships between economic benefits and energy consumption for busy servers and cooling system, energy price, and profit of the DC. These models define the problem of maximized economic benefit with the constraints.

Data Availability

The experimental data used to support the findings of this study are the dataset generated according to the strategy mentioned within the article by Arghode et al. [27]. The data generation programs can be easily implemented or are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank Li Jiyuan, who is a student of Thermal Engineering Department, School of Metallurgy, Northeastern University, for the calculation of the economic benefit model and improving the paper. This work was supported by the Research Grant from the National Natural Science Foundation of China (Grant no. 611662057) and the Fundamental Research Funds for the Central Universities (N182504017).

References

- [1] C. Nadjahi, H. Louahlia, and S. Lemasson, "A review of thermal management and innovative cooling strategies for data center," *Sustainable Computing: Informatics and Systems*, vol. 19, pp. 14–28, 2018.
- [2] S. Baig, W. Iqbal, J. Lluís Berral, and D. Carrera, "Adaptive sliding windows for improved estimation of data center resource utilization," *Future Generation Computer Systems*, vol. 104, pp. 212–224, 2020.
- [3] T. J. Chainer, M. D. Schultz, P. R. Parida, and M. A. Gaynes, "Improving data center energy efficiency with advanced thermal management," *IEEE Transactions on Components, Packaging and Manufacturing Technology*, vol. 7, no. 8, pp. 1228–1239, 2017.
- [4] T. L. Vasques, P. Moura, and A. De Almeida, "A review on energy efficiency and demand response with focus on small and medium data centers," *Energy Efficiency*, vol. 12, no. 5, pp. 1399–1428, 2019.
- [5] H. Rong, H. Zhang, S. Xiao, C. Li, and C. Hu, "Optimizing energy consumption for data centers," *Renewable and Sustainable Energy Reviews*, vol. 58, pp. 674–691, 2016.

- [6] P. Kumar and Y. Joshi, "Optimizing energy consumption for data centers Fundamentals of data center airflow management," *Energy Efficient Thermal Management of Data Centers*, Springer, Berlin, Germany, pp. 39–136, 2012.
- [7] W.-X. Chu, R. Wang, P.-H. Hsu, and C.-C. Wang, "Assessment on rack intake flowrate uniformity of data center with cold aisle containment configuration," *Journal of Building Engineering*, vol. 30, p. 101331, 2020.
- [8] J. Cho and J. Woo, "Development and experimental study of an independent row-based cooling system for improving thermal performance of a data center," *Applied Thermal Engineering*, vol. 169, p. 114857, 2020.
- [9] C. Jin, X. Bai, C. Yang, W. Mao, and X. Xu, "A review of power consumption models of servers in data centers," *Applied Energy*, vol. 265, p. 114806, 2020.
- [10] S. A. Nada and K. E. Elfeky, "Experimental investigations of thermal managements solutions in data centers buildings for different arrangements of cold aisles containments," *Journal of Building Engineering*, vol. 5, pp. 41–49, 2016.
- [11] M. Tatchell-Evans, N. Kapur, J. Summers, H. Thompson, and D. Oldham, "An experimental and theoretical investigation of the extent of bypass air within data centres employing aisle containment, and its impact on power consumption," *Applied Energy*, vol. 186, pp. 457–469, 2017.
- [12] V. Sundaralingam, V. K. Arghode, and Y. Joshi, "Experimental characterization of cold aisle containment for data centers," in *Proceedings of the 29th IEEE Semiconductor Thermal Measurement and Management Symposium*, March 2013.
- [13] C. Lyu, G. Chen, S. Ye, and Y. Liu, "Enclosed aisle effect on cooling efficiency in small scale data center," *Procedia Engineering*, vol. 205, pp. 3789–3796, 2017.
- [14] Y. Fulpagare, Y. Joshi, and A. Bhargav, "Rack level transient CFD modeling of data center," *International Journal of Numerical Methods for Heat & Fluid Flow*, vol. 28, no. 2, pp. 381–394, 2018.
- [15] K. Zhu, Z. Cui, Y. Wang, H. Li, X. Zhang, and C. Franke, "Estimating the maximum energy-saving potential based on IT load and IT load shifting," *Energy*, vol. 138, pp. 902–909, 2017.
- [16] S. K. Uzaman, A. u. R. Khan, J. Shuja, T. Maqsood, F. Rehman, and S. Mustafa, "A systems overview of commercial data centers," *International Journal of Information Technology and Web Engineering*, vol. 14, no. 1, pp. 42–65, 2019.
- [17] T. Y. Lawson and D. Zbigniew, "Economic framework for resource management in data centers," in *Proceedings of the 2016 IEEE International Conference on Communication Systems (ICCS)*, pp. 1–6, IEEE, Shenzhen, China, December 2016.
- [18] C. Gu, L. Fan, W. Wu, H. Huang, and X. Jia, "Greening cloud data centers in an economical way by energy trading with power grid," *Future Generation Computer Systems*, vol. 78, pp. 89–101, 2018.
- [19] W. Lu, P. Lu, Q. Sun, S. Yu, and Z. Zhu, "Profit-aware distributed online scheduling for data-oriented tasks in cloud datacenters," *IEEE Access*, vol. 6, pp. 15629–15642, 2018.
- [20] X.-D. Dong, S. Chen, L.-P. Zhao, X.-B. Zhou, H. Qi, and K.-Q. Li, "More requests, less cost: uncertain inter-datacenter traffic transmission with multi-tier pricing," *Journal of Computer Science and Technology*, vol. 33, no. 6, pp. 1152–1163, 2018.
- [21] R. Mokadem and A. Hameurlain, "A data replication strategy with tenant performance and provider economic profit guarantees in cloud data centers," *Journal of Systems and Software*, vol. 159, Article ID 110447, 2020.
- [22] S. M. MirhoseiniNejad, H. Moazamigoodarzi, G. Badawy, and G. Douglas, "Joint data center cooling and workload management: a thermal-aware approach," *Future Generation Computer Systems*, vol. 104, pp. 174–186, 2020.
- [23] K. Zhang, Y. Zhang, J. Liu, and X. Niu, "Recent advancements on thermal management and evaluation for data centers," *Applied Thermal Engineering*, vol. 142, pp. 215–231, 2018.
- [24] J. Ni, B. Jin, S. Ning, and X. Wang, "The numerical simulation of the airflow distribution and energy efficiency in data centers with three types of aisle layout," *Sustainability*, vol. 11, no. 18, 2019.
- [25] P. Dhoot, J. W. Vangilder, Z. Pardey, and C. M. Healey, "Zero-equation turbulence models for large electrical and electronics enclosure applications," US Patent 20180014427, 2018.
- [26] C. Chen and S. Jaw, *Fundamentals of Turbulence Modeling*, Taylor & Francis, Washington, DC, USA, 1998.
- [27] V. K. Arghode, V. Sundaralingam, and Y. Joshi, "Thermal characteristics of open and contained data center cold aisle," *Journal of Heat Transfer*, vol. 135, no. 6, 2013.
- [28] C. Da, P. Georges, and O. Wojciech, "Energy and thermal models for simulation of workload and resource management in computing systems," *Simulation Modeling Practice and Theory*, vol. 58, pp. 40–54, 2015.
- [29] A. M. T. Ramos, J. A. Carvalho, and G. L. Vasconcelos, "Exponential model for option prices: application to the Brazilian market," *Physica A: Statistical Mechanics and Its Applications*, vol. 445, pp. 161–168, 2016.
- [30] Z. Zhou, F. Liu, H. Jin, B. Li, B. Li, and H. Jiang, "On arbitrating the power-performance trade off in SaaS clouds," in *Proceedings of the 2013 Proceedings IEEE INFOCOM*, pp. 872–880, IEEE, Toronto, Canada, April 2013.