Abstract

Data center is a fundamental infrastructure of computers and networking equipment to collect, store, process, and distribute huge amounts of data for a variety of applications such as Cyber–Physical–Social Systems, business enterprises and social networking. As the demands of remote data services keep increasing, both the workload of the data center and its [power consumption](https://www.sciencedirect.com/topics/engineering/electric-power-utilization) are rapidly rising. An indispensable part of a data center is the cooling system which provides a suitable operation environment, and accounts for around 30% of the [power consumption](https://www.sciencedirect.com/topics/engineering/electric-power-utilization) of the data center. Therefore, optimized [energy management](https://www.sciencedirect.com/topics/engineering/energy-management) of data center’s cooling system is a highly profitable research area. Generally, a cooling system is made up of a [mechanical refrigeration](https://www.sciencedirect.com/topics/engineering/mechanical-refrigeration) sub-system and a terminal cooling sub-system. Heat generated during operation of the data center will be absorbed by the latter one, and transferred into the outdoor environment via the former one. Depending on the cooling principle, current cooling solutions can be classified into air-cooling, liquid-cooling or free cooling technology. Although air-cooling is widely used in most existing data centers, the other two solutions have attracted more interests due to their excellent cooling effectiveness and higher energy efficiencies. Among the different cooling equipment, the [chillers](https://www.sciencedirect.com/topics/engineering/chiller) and fans are the major power consumers of the entire cooling system. Therefore, modeling of their power consumption is important for [energy management](https://www.sciencedirect.com/topics/engineering/energy-management) of the cooling system, which can be classified into mechanism-based methods and data-driven methods. Based on the aforementioned models, [optimization strategies](https://www.sciencedirect.com/topics/engineering/optimization-strategy) for the operation management of cooling equipment are proposed to reduce the power consumption of the cooling system, which mainly includes the model predictive control-based methods and reinforcement learning-based methods. This paper is an overview of the data center’s cooling system, which mainly includes the mainstream cooling solutions, the power [consumption modeling](https://www.sciencedirect.com/topics/engineering/consumption-modeling) methods and the optimization control strategies. In addition, several current challenges and future work in the data center’s cooling system are described.

Introduction

Nowadays, data center (DC) has become a fundamental infrastructure of people’s daily lives, providing remote data services for business, entertainment and many other human’s requirements. With centralized computation and storage resources (CCSRs), DCs can resolve the issue of large-scale data processing and storage simultaneously for massive users [1], [2], [3]. Generally, the physical components of a DC include information and communications technology (ICT) devices and auxiliary equipment. The ICT devices are in charge of data center’s pivotal functions such as data transmission and processing, including the servers, the network switches and the routers. The auxiliary equipment, such as the cooling system and the power supply system, are used to ensure the stable operation of the ICT devices. For the ICT devices, the operation temperature is an important factor that can greatly affect the stability of their performance. In the era of big data, the booming demands for cloud servers greatly facilitate the expanding scales of DCs [1], [3], which also keeps the DC’s workloads increasing. However, more ICT devices under high computational loads will generate a great amount of heat, which can seriously affect their stable operation. To provide a suitable environment for them, the cooling system has become an indispensable part of a DC, which also brings high power consumption. Therefore, the heat removal by the cooling system is one of the most prominent challenges in the maintenance of DCs [2], [4], [5].

In general, a typical DC cooling system is made up of a mechanical refrigeration sub-system (MRSS) and a terminal cooling sub-system (TCSS) [6]. The MRSS, usually comprised of equipment such as chillers, pumps and cooling towers, provides cold supply for the heat removed in the TCSS. The TCSS transfers heat from the indoor environment to the outdoor environment via the MRSS, the main solutions of which are air-cooling, liquid-cooling and free-cooling techniques. During DC’s operation, the cooling system accounts for around 30% of the total power consumption. According to Fig. 1, the main power consumers of the cooling system are the fans of the computer room air conditioner (CRAC) and the chiller. To reduce their power consumption, adaptive control strategies using different algorithms have been proposed to adjust their set-points by real-time conditions such as workloads and operation temperatures. As two main paradigms to obtain the optimal control strategy, model-predictive control (MPC) [7], [8], [9] based on system dynamics modeling of cooling equipment, or reinforcement learning control (RLC) [10], [11], [12], [13] analyzing the system feedbacks of decisions, can be used. Given the escalating workloads of DC, research on the operation management of the cooling system should not only focus on the cooling technique itself, but also consider the control strategies with the characteristics of the scenarios and the users in the DC data services.

In recent years, the powerful CCSRs provided by DCs have facilitated the development of Cyber–Physical–Social Systems (CPSSs) [14]. A CPSS, representing a holistic architecture with data collection devices, data transmission networks and system actuators, is applied to establish interactions among cyberspace, the physical environment and human society such as in smart manufacturing [15], smart city [16], wearable sensors [17], [18], [19], smart homes [20], [21], [22], and social dispersed computing [23]. In CPSS applications, multi-sources heterogeneous data is acquired from human, equipment objects and sensors in the social and physical layers, which becomes continuous dataflow at the edge of Internet-of-Things (IoT) networks [24]. The DC’s central cloud continuously receives a huge amount of data, processing and analyzing it to provide information cognition and feedback execution to the edge data sources. Recently, under the impact of COVID-19, the lockdown significantly increased the business demands of cloud-based CPSS deployment for its stable remote access [25], which also bring challenges to DC’s auxiliary equipment especially the cooling system.

As IoT technology is introduced in a DC’s maintenance, the cooling system is equipped with a cyber layer for status monitoring and remote control, thus becoming a cyber–physical system. Since the Human-in-the-Loop (HitL) also exists in a DC cooling system, it will be beneficial for its power consumption efficiency enhancement to take the social factors into consideration. The benefits can be summarized by the following three points, which is illustrated in Fig. 2. **(1) *Considering Human’s activity impacts on a DC cooling system.*** In a specific application scenario, the workloads of ICT devices in different time periods are unbalanced, which makes the required cooling capacity in each period different. For example, the COVID-19 outbreak has made online meetings common in business scenarios. Higher workloads of a DC by the transmission and processing of heterogeneous data such as text and video information, which is increased significantly during people’s working hours. The cooling system should also be deployed with higher cooling power in those time periods, to keep the ICT devices stable operations under heavy workloads and reduce the power consumption cost in other free periods. **(2) *Effective Human–Computer interactions.*** It makes the DC monitoring system more efficient if it can be embedded with reliable human–computer interactions. Technicians are able to remotely acquire the real-time operating status such operation power of different devices such as cooling equipment. They can also carry out real-time control to different equipment, and receive successful messages when their commands take effect. **(3) *Intelligent Human Strategies Modeling.*** With different ICT workloads and specific environmental conditions, the corresponding adjustments made by technicians to cooling equipment can be valuable expert data. Artificial intelligence methods can be used to learn the control experience from the analysis of such data, which can build the intelligent control models to adaptively regulate the system under different conditions to reduce the power consumption.

To improve the power consumption efficiency of DC cooling systems, hardware optimization, such as the layout design and cooling equipment upgrade of TCSS and MRSS, is a basic aspect. Given the impacts by CPSS characteristics, it is also important to give the cooling system the capabilities of adaptive workload regulation and intelligent control scheduling. Therefore, this survey mainly discusses the classical solutions and recent development of the software and hardware components in DC cooling systems. The contributions can be mainly summarized into the following points. **(1) *Summarizing Thermal Efficiency Enhancement Technologies of DC Cooling hardware.*** The configurations of typical cooling systems in DCs, including air-cooling, liquid-cooling or free-cooling, will be introduced in Sections 2 Air-cooling technology, 3 Liquid-cooling technology, 4 Free cooling technology. For each of them, their unique hardware designs and the characteristics of their MRSS and TCSS are discussed in detail. By analyzing their advantages and limitations, the suitable scenarios of these three techniques are also summarized. **(2) *Analyzing State-of-the-art Software Development on DC Cooling Power Modeling and Optimization Control.*** The power consumption modeling of the two major power consumers, the chiller and the CRAC fans, are presented in Section 5.1. The mechanism-based modeling methods of the CRAC fans and the latest data-driven methods for the chillers are mainly introduced. Also, their merits and drawbacks are discussed. The MPC and RLC-based methods, as two mainstream adaptive control strategies, are introduced in Section 5.2 from the latest publications. The principles of the two algorithms and how they are used for DC’s cooling control are presented. Moreover, the similarities and differences between MPC and RLC are discussed. **(3) *Summarizing Promising Directions for Current and Future Development of DC Cooling.*** The challenges which currently affect the DC cooling performances are summarized with their possible solutions, in Section 6.1. How emerging technologies can be used to improve DC cooling operation in future are also included in Section 6.2.

Section snippets

Air-cooling technology

Air-cooling technology is the most conventional solution that is widely applied in large scale data centers. Its main advantages are simple maintenance and acceptable operation cost [2], [39]. The basic mechanism of the air cooling technology is illustrated in Fig. 3. It is accomplished by the airflow cycle in TCSS, whose cold supply is provided by the computer room air conditioner (CRAC). In the next two subsections, we will discuss terminal cooling and mechanical refrigerating air cooling

Liquid-cooling technology

Air-cooling is the most common cooling method used in data centers. However, it is an inefficient cooling method due to the low density and heat dissipation capacity of air. In contrast, liquid-cooling is one of the most effective methods, saving total energy consumed compared to air-cooling system. According to how liquid coolant contacts heat source, liquid-cooling methods can be mainly divided into indirect and direct liquid-cooling methods. Similar with air-cooling methods, liquid-cooling

Free cooling technology

By leveraging natural cold source, free cooling can reduce the overall energy consumption of data centers. As summarized in Refs. [66], [67], [68], free cooling technologies can be commonly classified as air-side, water-side and heat pipe based systems. As the names suggest, this classification method derives from the classical technologies described in Sections 2 Air-cooling technology, 3 Liquid-cooling technology . To explicitly present the optimization strategies, we summarize free cooling

Power consumption modeling and optimization methods

In the MRSS of a data center, much power will be consumed for the operations of cooling equipment. Therefore, optimization methods are needed to provide set-points for cooling devices to reduce the power consumption of their operations. This will be introduced in Section 5.1. Power consumption modeling using relations between the cooling equipment measurable variables and the power consumption for the optimization methods will be introduced in Section 5.2.

Current challenges and future work

Based on the discussions of technology, power consumption modeling and control strategy optimization of the cooling system of a data center, several challenges are identified. Some current and future challenges are shown in Fig. 11.To focus our discussion, we select three important current challenges and present possible solutions. Also, to build a cooling system with higher energy efficiency for the CPSS’s data center, some additional considerations should be investigated and implemented. The

Conclusions

A data center is the fundamental infrastructure in the era of big data. The cooling system of a data center is an indispensable component to provide a suitable operation temperature for the information and communications technology devices. This paper presents an survey of the technology, power consumption modeling and control strategy optimization of a data center’s cooling system. Current cooling solutions are classified into air-cooling, liquid-cooling and free cooling technology. The

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Qingxia Zhang** is a Ph.D candidate of school of Computer Science, Fudan University. She received the M.S. degree in School of Computer Science & Technology from Northeastern University, Shenyang, China, in 1997. Her research interests are computer technology, big data analysis and natural language processing.