CHEUNG, Howard; WANG, Shengwei. Reliability and availability assessment and enhancement of water-cooled multi-chiller cooling systems for data centers. Reliability Engineering & System Safety, 2019, 191: 106573.

https://doi.org/10.1016/j.ress.2019.106573

Reliability and availability assessment and enhancement of water-cooled multi-chiller cooling systems for data centers

HOWARD CHEUNG, SHENGWEI WANG

Department of Building Services Engineering, The Hong Kong Polytechnic University, Hong Kong

Keywords*: data center; cooling; reliability; availability; chiller*

Abstract

There are strict requirements and needs to avoid huge downtime cost for data center cooling systems to operate continuously for reliable data center services. Current design standards achieve this by installing redundant cooling equipment. Other equipment including distribution headers is also used. However, *no study is found to quantify the reliability of these designs comprehensively* using on-site performance data. It is unknown how much reliability can be improved by adopting these designs or if they are burdens to the cost of systems only. This study quantifies how different data center cooling system configurations enhance the reliability and availability of data centers. The results show that it is crucial to install a redundant chiller or redundant chiller plant with alternative cooling sources to meet the requirement of high-tier data centers, but other common practices such as distribution headers and the 2(N+1) configuration do not improve the reliability and availability of data center cooling systems effectively.

1. Introduction

Data center operation downtime is becoming a growing concern among many companies to avoid heavy down-time cost (e.g. USD$5,600 per minute in 2011 [52]) and termination of important services such as devices in smart homes, banking or even government bodies [1]. To reduce the operation downtime, data centers are always designed for high reliability. Turner et al. classified 4 tiers of data centers in *Data center site infrastructure tier standard: Topology* in accordance with their reliability as shown in Table 1 [2], [3].

Table 1: Description and downtime of data centers of different tiers

Tier	Description	Annual end-user downtime	Resultant end-user availability	Number of active components to support load	Distribution paths
I	Basic site infrastructure	28.8 hours	99.67%	N	1
II	Redundant site infrastructure capacity components	22.0 hours	99.75%	N+1	1
III	Concurrently maintainable site infrastructure	1.6 hours	99.98%	N+1	1 Active and 1 Alternate
IV	Fault tolerant site infrastructure	0.8 hours	99.99%	N after any failure	2 Simultaneously active
Remark: N is the number of components required at maximum load (i.e. number of duty equipment).					

Table 1 shows that the higher the tier, the shorter its downtime. To build a reliable data center, the current standard recommends designing with redundancy, designing for concurrent data center operation and maintenance and embedding fault tolerant design into the data center. Since data centers require continuous cooling to maintain service (e.g. disruption of data center cooling leads to rapid increase of data center temperature [55] and is one of the four major causes of extremely serious outages [56]), data cooling system design also needs to follow the guideline in Table 1. Research is being done to enhance data center reliability by various methods such as installing redundant equipment [3]–[5], using power storage systems (e.g. uninterruptible power supply (UPS)) [6], [7], power management [8], [9], workload management [10], [11], fault detection and diagnostic algorithms [12]–[14], fault-tolerant thermal management [15] and data center network management [16]–[19]. Using multiple data centers in various locations to enhance the overall reliability of data center services of a company is also investigated though the solution does not enhance the reliability of any data center [1], [20]–[22].

Conventionally, data centers follow the tier classification scheme to ensure that their reliability reaches the required level [2], [3]. Their designs follow the requirements by adding redundant power supply and cooling system and energy distribution paths. UPS systems are added to the electricity distribution systems of the IT equipment, the ventilation equipment or even the cooling systems to ensure that both electricity and cooling can be provided continuously in the event of maintenance or faults [2], [23]. While Tier I data centers in Table 1 do not require additional equipment, Tier II data centers require installation of some extra equipment such as UPS systems and an emergency generator. Tier III data centers require more redundant equipment, and they need alternative power and cooling distribution systems. Tier IV data centers require more equipment to avoid total failure, including fault tolerant equipment, continuous operation of two different power and cooling supply systems and extra UPS systems for all operating systems. Some of them may be built with 2N power infrastructure and N+1 cooling equipment to satisfy the requirement [24], while others may contain 2(N+1) cooling systems to maintain the reliability [25]. The design method of data centers with redundancy plays a critical role in the reliability of data center operation.

While redundancy guarantees an increase of reliability, it is not easy to add redundancy as the cooling technology of data centers becomes more complicated for higher energy efficiency. For example, water-cooled chiller plants and economizers are used to replace air-cooled chillers due to its higher efficiency [26], [27], but their installation requires much more space, structural support and extra utility services such as cooling water pumps, water storage boxes and water utility. It is more difficult to install redundant equipment for these setups. The use of combined cooling, heating and power (CCHP) plant in data centers requires even more equipment such as gas-powered generators and absorption chillers [28], [29]. The complexity of the new cooling technology makes it much more difficult to add redundant equipment, and it is necessary to have a quantitative understanding of the benefit of the redundancy to the reliability of a data center.

Although it is well understood that redundant power infrastructure and cooling equipment improve the reliability of a data center, the benefit of the redundancy to the reliability of data center operation is seldom quantified. Wiboonrat calculated the availability of components and the overall data center by simulation and used a reliability block diagram to recommend redundant power links to enhance a Tier IV data center availability well above the requirement in Table 1 [30]. Nguyen conducted a similar analysis on the reliability and availability of cloud data center networks using hierarchical models, but the study did not consider the details within data center cooling systems such as the difference of chillers between different types of cooling systems [57]. Gomes et al. used Petri Nets to model the availability of data centers under Tier I and II architecture but did not quantify the effect of redundancy at higher tier classes [58]. Callou et al. developed a cost model of data center cooling system based on the its energy consumption and reliability and tested the model with redundant equipment, but the study did not quantify the effect of a redundant chiller plant [59]. Stein looked at studies on two data centers and concluded that the use of air-side economizers did not reduce the reliability of a data center [31]. Wang et al. studied how different operation modes of water-side economizers affect the reliability of a data center [54]. Multiple studies have been conducted on the control of redundant computer room air conditioners (CRACs) or fans to fulfill the temperature requirement inside a data center without failing any computing equipment [15], [32]–[34]. There were also studies that proposed control algorithms to reduce the use of redundant CRACs to enhance data center efficiency [35], [36]. Wang et al. found that redundant cooling systems in a building helped to increase the availability of a cooling system from 98.50% to 99.99% [37]. Similar techniques are used to optimize a combined cooling and heat system by reliability and energy efficiency [38]. Gang et al. demonstrated how to design a cooling system that meets its reliability requirement using Monte-Carlo-Markov-Chain method [39]. Wiboonrat simulated the reliability of data centers with 2(N+1) CRACs and described a method to manage a data center based on reliability by separating the data center components into multiple zones [40]. Despite the various studies on data center reliability and the emphasis of various studies to use different redundancy levels such as N+1 and 2(N+1) cooling systems to enhance reliability [24], [25], *few studies quantified the improvement of the reliability of a cooling system brought by equipment redundancy levels commonly used in data center designs.*

Besides using redundant cooling equipment to enhance cooling system reliability, the studies also ignored the effect of the use of distribution headers in chiller plants to enhance the reliability of a water-cooled data center cooling system. Unlike systems using CRACs that operate with their own compressors and cool data centers by generating cold air using their own vapor compression systems, large-scale water-cooled data center cooling systems use CRAHs and chillers. CRAHs cooled data centers by transferring heat from data centers to chilled water generated by chillers. For systems with CRAHs and chillers, there are always claims that they can use distribution headers to enhance its reliability by enabling more available equipment in a cooling system as shown in Figure 1 and Figure 2.

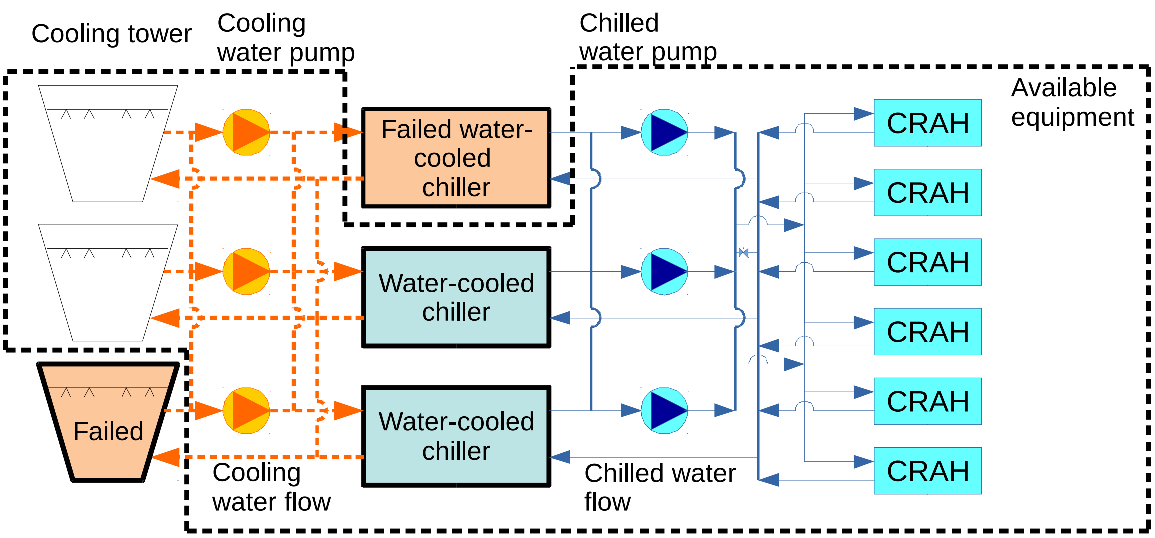

Figure 1 A cooling system with distribution headers

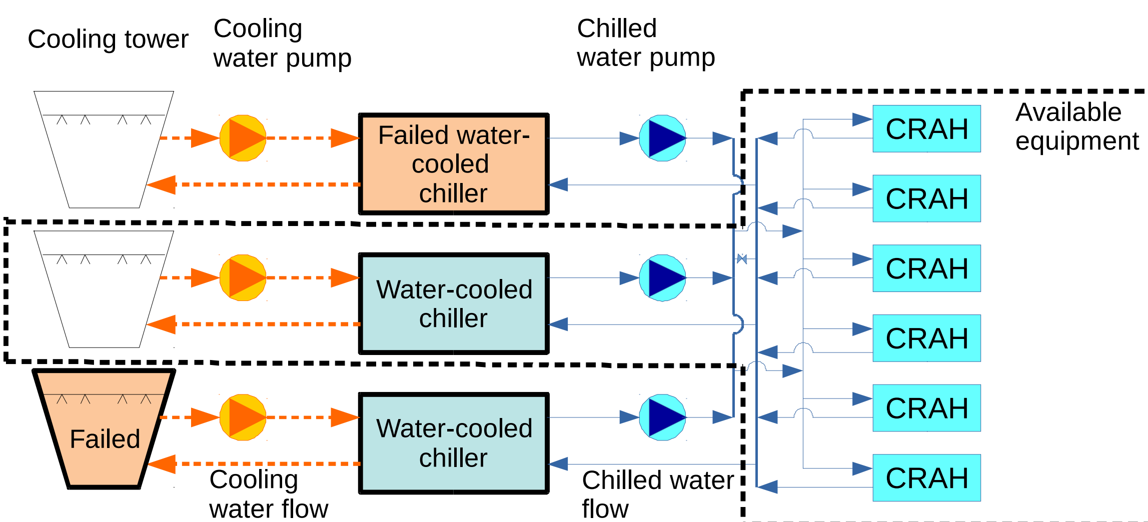

Figure 2 A cooling system without distribution headers

Some have studied the benefits of the use of distribution headers to reduce energy use of a water-cooled data center. By using distribution headers, one can operate fewer pumps than the number of operating chillers under some specific operation conditions and reduce the pump power consumption [41]. One can also operate more cooling towers than necessary to increase the evaporatively cooling effectiveness for higher overall energy efficiency [48]. However, there are no studies on the quantification of the benefit of distribution headers to the reliability of a cooling system. Although the distribution headers may increase the reliability as much as redundant equipment, it is unknown how much reliability they can bring to the system quantitatively. Some current designs of cooling systems still contain no distribution headers among equipment to facilitate simple control algorithms [42], [43]. *It is unknown if distribution headers in a cooling system that are much cheaper than extra equipment can be used to replace redundant equipment for enhancing the reliability of a data center. In fact, the reliability enhancement of common data center cooling system designs is seldom quantified. Designs for reliability of cooling systems are mostly determined by engineers subjectively, and it is unknown if the designs can really enhance the system reliability to the required level or is a burden to the data center construction and operation cost only.*

This study aims to quantify benefits of these design methods on the reliability and availability and determine their necessity to the availability required for different tiers of systems. Reliability and availability analyses are conducted to assess and quantify the benefit of redundant equipment and distribution headers to the reliability and availability of water-cooled data center cooling systems and identify which types of cooling system designs are preferable for highly reliable data centers. It uses existing theories on system reliability and availability and on-site observations of equipment reliability to quantify the reliability and availability of a cooling system under different configurations. It compares the reliability of the scenarios built under the configurations and concludes if common design methods of cooling systems are necessary for a highly reliable data center.

2. Theory to quantify reliability and availability of a water-cooled data center cooling system

To quantify the reliability and availability of a water-cooled data center cooling system with distribution headers and redundant equipment, it is necessary to understand the quantification of the reliability and availability of an equipment, the reliability and availability of groups of equipment connected in series by water pipes, the reliability and availability of equipment connected in parallel by water pipes, and the reliability and availability of a general data center cooling system with chiller plants and CRAHs.

2.1 Reliability and availability of an equipment

Reliability of an equipment is the probability which an equipment continues to operate without failure after operation for a certain period. It can be quantified using Equation (1).

$$R(t) = \exp\left(-\lambda t\right) \tag{1}$$

where R is the reliability of an equipment, t is the operation time and λ is the failure rate of the equipment.

While the reliability of an equipment is a function of time, its failure rate λ is usually assumed to be time-invariant and thus can be used to represent an equipment's reliability. The lower the failure rate of an equipment, the longer the equipment operates without failure. It is usually calculated with the repair rate as shown in Equations (2) and (3).

$$\lambda = \frac{1}{MTTF} \tag{2}$$

$$\mu = \frac{1}{MTTR} \tag{3}$$

Mean time to failure (MTTF) is the average time that an equipment fails as it operates normally, and mean time to repair (MTTR) is the average time that an equipment is repaired from its failed state [39]. The failure rate and repair rate can be used to calculate the availability of an equipment. Availability of an equipment is the proportion of time that an equipment is available for operation. It is also the probability of an equipment that is available to operate and is used to define the tier classification of data centers [1]. It is related to the probability matrix of the state of an equipment. The matrix includes the probabilities of an equipment to transition or to remain in different states. For availability studies, the states involved for an equipment are normal state and failed state. When an equipment is normal, there is a probability for it to stay normal and there is also a probability for it to fail. When a machine has failed, there is also a probability for it to remain failed and a probability for it to be repaired to its normal state. Hence there are four probabilities in a probability matrix of one equipment to describe the transition between states of an equipment as shown in Equation (4).

$$P = \begin{bmatrix} P_{nn} & P_{nf} \\ P_{fn} & P_{ff} \end{bmatrix} \tag{4}$$

where **P** is the probability matrix, P_{nn} is the probability of an equipment to remain normal, P_{nf} is the probability of an equipment to fail from its normal state, P_{fn} is the probability of an equipment to be repaired from its failure, and P_{ff} is the probability of an equipment to remain failed.

Assuming that the failure rate and repair rate only depends on the state of an equipment, the probability matrix of the states of an equipment after an operation or maintenance time *t* from the last state transition can be quantified by Equation (5) using the constant failure rate model [44].

$$P(t) = \begin{bmatrix} exp(-\lambda t) & 1 - exp(-\lambda t) \\ 1 - exp(-\mu t) & exp(-\mu t) \end{bmatrix} \quad (5)$$

where $exp(-\lambda t)$ is the reliability function of the equipment as shown in Equation (1)

However, to use Equation (5), the operation time of an equipment under normal state and the maintenance time of a machine is needed, and the calculation of the matrix requires computational expensive dynamic simulation of an equipment. To simplify the calculation, the steady state assumption used for the reliability and availability calculation of chiller plants in previous studies is used [39], [45], and steady-state probability of an equipment under the normal state of an equipment is calculated by Lisnianski et al. [46]. The stead-state probability of an equipment to be in its normal state is the availability of an equipment, and the equation of the probability of normal state is shown in Equation (6).

$$p_{nor} = \frac{\mu}{\mu + \lambda} \quad (6)$$

where p_{nor} is the probability of an equipment to be at its normal state.

In cases which only the availability and failure rate of an equipment are known, the repair rate can also be calculated as a function of both variables as shown in Equation (7).

$$\mu(\lambda, p_{nor}) = \frac{\lambda p_{nor}}{1 - p_{nor}} \quad (7)$$

2.2 Reliability and availability of a group of equipment connected in series by water pipes

For equipment that are connected in series by water pipes, all of them become unavailable for cooling operation when any of them fails. Their failure rate is the sum of the failure rates of all equipment connected in series as shown in Equation (8).

$$\lambda_{series} = \sum_{i=1}^{n_{series}} \lambda_i \quad (8)$$

where λ_{series} is the overall failure rate of a group of equipment connected in series, n_{series} is the number of equipment connected in series, and λ_i is the failure rate of the i^th^ equipment in the group.

Their overall availability is the product of their availabilities as shown in Equation (9).

$$p_{nor,series} = \prod_{i=1}^{n_{series}} p_{nor,i} \tag{9}$$

where $p_{nor,series}$ is the availability of a group of equipment connected in series, and $p_{nor,i}$ is the availability of the ith equipment in the group.

2.3 Reliability and availability of a set of equipment connected in parallel by distribution headers

When a chiller plant equipment with distribution headers around it encounters a failure such as the failed chiller in Figure 1, because there are idle equipment of the same type connected to the same distribution headers, the idle equipment can be switched on to replace the failed equipment without any service interruption. Their overall availability is given by the availability of having at least *m* equipment that is in its normal state given a total number of equipment *M* surrounded by the two distribution headers. By binomial distribution [47], the probability is given by Equation (10) and (11).

$$p_{nor,par}(m, M, p_{nor}) = \sum_{k=m}^{M} p_{nor,par,only}(k, M, p_{nor}) \tag{10}$$

$$p_{nor,par,only}(k, M, p_{nor}) = \binom{M}{k} p_{nor}^{i} (1 - p_{nor})^{(M-i)} \tag{11}$$

where $p_{nor,par}$ is the availability of a group of equipment connected in parallel, $p_{nor,par,only}$ is the availability of a group of equipment when only k number of equipment is available, *m* is the number of required equipment in the group according to the data center controller, and *M* is the total number of equipment in the group.

By assuming that the failure rate of an equipment is much greater than that of the repair rate, the overall failure rate of the group of equipment, which is the failure rate of the *m* required equipment among *M* equipment in the group, can also be approximated as shown in Equations (12) [53].

$$\lambda_{par}(m, M, \lambda, \mu) = m\lambda \binom{M}{m} \left(\frac{\lambda}{\mu}\right)^{M-m} \quad (12)$$

where λ_{par} is the overall failure rate of the group of *M* equipment when *m* of them is needed.

2.4 Reliability and availability of a data center cooling system with multiple chiller plants

The failure rate and the availability of a data center cooling system using multiple identical chiller plants can be calculated using the configuration in Figure 3.

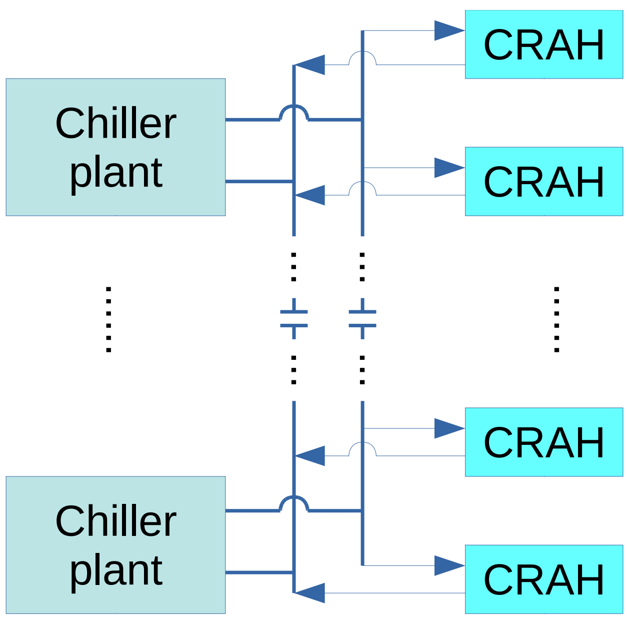

Figure 3: A generalized data center cooling system with CRAHs and chiller plants

The figure shows a data center cooling system supported by groups of chiller plants and CRAHs. Each chiller plant is connected to a group of CRAHs. A chiller plant and its

CRAHs can support the maximum cooling load of the data center when all its equipment can operate normally, and other chiller plants are all redundant chiller plants. The chiller plants have separate chilled water paths to satisfy the requirement to have more than 1 distribution path in Tier III and IV data centers in Table 1. This enables the formulation of the equations for the availability of the data center cooling system in Figure 3 using Equations (9) and (10). The resultant equations are Equations (13) and (14).

$$p_{sys,nor} = p_{nor,par}\left(m_{plant}, M_{plant}, p_{plant,nor}p_{CRAH,gp,nor}(m_{CRAH}, M_{CRAH})\right) \tag{13}$$

$$p_{CRAH,gp,nor}(m_{CRAH}, M_{CRAH}) = p_{nor,par}\left(m_{CRAH}, M_{CRAH}, p_{CRAH,nor}\right) \tag{14}$$

where $p_{sys,nor}$ is the availability of a data center cooling system, m_{plant} is the number of required chiller plant according to the data center control, M_{plant} is the total number of chiller plants in the data center, $p_{plant,nor}$ is the availability of a chiller plant, $p_{CRAH,gp,nor}$ is the availability of a group of CRAH connected in parallel, m_{CRAH} is the number of required CRAH to support the data center operation, M_{CRAH} is the number of CRAHs connected to a chiller plant, and $p_{CRAH,nor}$ is the availability of a CRAH.

The formulation of the equations for the failure rate of the data center cooling system in Figure 3 can be done using Equations (7) (8) and (12). The resultant equations are Equations (15), (16) and (17).

$$\lambda_{sys} = \lambda_{par}\left(m_{plant}, M_{plant}, \lambda_{plant} + \lambda_{CRAH,gp}(m_{CRAH}, M_{CRAH}), \mu_{plant+CRAH,gp}(m_{CRAH}, M_{CRAH})\right) \tag{15}$$

$$\mu_{plant+CRAH,gp}(m_{CRAH}, M_{CRAH}) = \mu\left(\lambda_{plant} + \lambda_{CRAH,gp}(m_{CRAH}, M_{CRAH}), p_{plant,nor}p_{CRAH,gp,nor}(m_{CRAH}, M_{CRAH})\right) \tag{16}$$

$$\lambda_{CRAH,gp}(m_{CRAH}, M_{CRAH}) = \lambda_{par}(m_{CRAH}, M_{CRAH}, \lambda_{CRAH}, \mu_{CRAH}) \tag{17}$$

where $\lambda_{sys,nor}$ is the failure rate of a data center cooling system, $\lambda_{plant,nor}$ is the failure rate of a chiller plant, $\lambda_{CRAH,gp,nor}$ is the failure rate of a group of CRAH connected in parallel, and $\lambda_{CRAH,nor}$ is the failure rate of a CRAH.

3. Description of cooling system configurations under study and the assessment procedure

To examine the effect of redundant equipment and distribution headers to the reliability of a data center cooling system, this study creates a baseline scenario of a data center cooling system and calculates the changes of the reliability of the system when different configurations and operating conditions are applied to the baseline scenario. These configurations include various levels of redundant equipment and distribution headers. The typical availability of individual equipment such as water-cooled chillers are also needed to conduct the calculation.

3.1 Baseline scenario

The baseline scenario contains 8 CRAHs and 1 water-cooled chiller plant as shown in in Figure 4. Water-cooled chillers are used in the baseline scenario because they are commonly recognized as the energy-saving designs relative to air-cooled setups [26]. The chiller plant has four chillers supporting a data center with N = 4, and it is designed to operate all chillers when the data center cooling load reaches its maximum. Its single-speed cooling water pumps, variable-speed chilled water pumps and its cooling towers are dedicated to different chillers. The chilled water from the chillers are delivered to CRAHs, and each CRAH uses a blower and a heat exchanger to absorb heat from the hot air generated by the data center IT equipment to the chilled water to keep the data center cool enough for continuous operation. While pumps are dedicated to chillers, distribution headers around the CRAHs are typical setups in the design of cooling systems [42], [48].

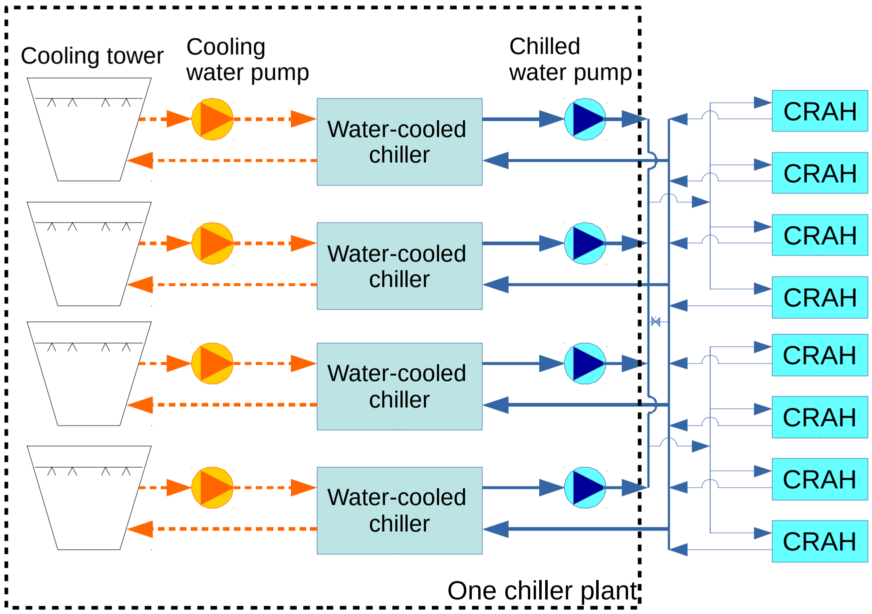

Figure 4: Configuration in the baseline scenario

To calculate the failure rate and the availability of the baseline scenario, the number of equipment required in the chiller plant to support the data center operation must be defined. When one chiller is needed to generate enough cooling to satisfy the cooling load of a data center, it is assumed that one cooling tower, one cooling water pump, one chilled water pump and two CRAHs connected to the chiller are needed to deliver the cooling to the data center and to reject the waste heat to the ambient. When the cooling load increases and more chillers are needed to satisfy the cooling demand, the numbers of operating cooling towers, cooling water pumps, chilled water pumps and CRAHs increase proportionally. The equations to calculate the availability of the baseline scenario can be derived using Equations (9) and (10), and the resultant equations are Equations (18), (19) and (20).

$$p_{sys,base,nor} = p_{wplant,base,nor} p_{CRAH,gp,nor}(2m_{ch}, 2N) \quad (18)$$

$$p_{wplant,base,nor} = p_{water} p_{ctrl} p_{nor,par}\left(m_{ch}, N, p_{cdp+wch+chp+ct,nor}\right) \quad (19)$$

$$p_{cdp+wch+chp+ct,nor} = p_{cdp,nor} p_{wch,nor} p_{chp,nor} p_{ct,nor} \quad (20)$$

where $p_{sys,base,nor}$ is the availability of a data center cooling system in the baseline scenario, $p_{wplant,base,nor}$ is the availability of a water-cooled chiller plant in the baseline scenario, p_{water} is the availability of water utility, p_{ctrl} is the availability of the chiller plant control box, m_{ch} is the number of required chiller, $p_{cdp,nor}$ is the availability of a single-speed cooling water pump, $p_{wch,nor}$ is the availability of a water-cooled chiller, $p_{chp,nor}$ is the availability of a chilled water pump, $p_{ct,nor}$ is the availability of a cooling tower.

Equation (18) describes the availability of the data center cooling system under different number of required chillers in the baseline scenario. In this scenario, there are four chillers in its chiller plant, and each chiller is connected in series with a cooling water pump, a chilled water pump and a cooling tower. The chiller plant is connected to a group of CRAHs in series to form the entire data center cooling system in the baseline scenarios. These CRAHs are connected to each other in parallel using distribution headers. Thus Equation (19) is used to calculate the availability of the chiller plant, and Equation (20) is used to calculate the availability of the group of CRAHs.

The overall failure rate at different number of required chillers can then be given by Equations (21), (22), (23) and (24).

$$\lambda_{sys,base} = \lambda_{wplant,base} + \lambda_{CRAH,gp}(2m_{ch}, 2N) \tag{21}$$

$$\lambda_{wplant,base} = \lambda_{water} + \lambda_{ctrl} + \lambda_{par}\left(m_{ch}, N, \lambda_{cdp+wch+chp+ct}, \mu_{cdp+wch+chp+ct}\right) \tag{22}$$

$$\mu_{cdp+wch+chp+ct} = \mu\left(\lambda_{cdp+wch+chp+ct}, p_{cdp+wch+chp+ct,nor}\right) \tag{23}$$

$$\lambda_{cdp+wch+chp+ct} = \lambda_{cdp} + \lambda_{wch} + \lambda_{chp} + \lambda_{ct} \tag{24}$$

where $\lambda_{sys,base}$ is the failure rate of a data center cooling system in the baseline scenario, $\lambda_{wplant,base}$ is the failure rate of a water-cooled chiller plant in the baseline scenario, λ_{water} is the failure rate of water utility, λ_{ctrl} is the failure rate of the chiller plant control box, m_{ch} is the number of required chiller, λ_{cdp} is the failure rate of a single-speed cooling water pump,

λ_{wch} is the failure rate of a water-cooled chiller, λ_{chp} is the failure rate of a chilled water pump, and λ_{ct} is the failure rate of a cooling tower.

3.2 Different cooling load as different operating conditions

Because the number of required operating equipment changes the failure rate and the availability of a data center, operating conditions corresponding to different cooling load are set up using different number of required equipment according to Table 2 to examine how cooling load affects the availability of a scenario.

Table 2: Number of operating equipment under different cooling load and hence different number of required equipment

Number of required equipment	Number of operating equipment				
	Chillers	Cooling towers	Cooling water pumps	Chilled water pumps	CRAHs
N-3	1	1	1	1	2
N-2	2	2	2	2	4
N-1	3	3	3	3	6
N	4	4	4	4	8

When the cooling load is low, not all chillers are needed. This is modeled by the N-3 and N-2 cases in Table 2. When the cooling load is high, almost all chillers are operated to satisfy the cooling load. This is modeled by the N-1 and N cases in Table 2.

3.3 Configurations with redundant equipment

The baseline scenario assumes that all equipment must be used to satisfy the maximum data center cooling load. Unlike the baseline scenario, high-tier data centers in Table 1 have redundant equipment to enhance the reliability and availability of its cooling system. They have more than N equipment and even more than 1 chiller plant to reduce their system downtime. The redundant equipment is also called stand-by equipment. The reliability of cooling systems in these data centers are considered in this study using the configurations in Table 3.

Table 3: Levels of redundant equipment under study

Total number of equipment	Redundant equipment as an extra chiller plant	Number of equipment contained in the extra plant	Chillers in the extra plant being replaced by air-cooled chillers	Redundant controller for the chiller plant
N+1	No	N/A	No	No
N+2	No	N/A	No	No
2N	Yes	N	No	No
2N	Yes	N	Yes	No
2(N+1)	Yes	N+1	No	No
2(N+1)	Yes	N+1	Yes	No
2N	No	N/A	No	Yes
2(N+1)	No	N/A	No	Yes

In Table 3, there are four levels of extra equipment. They correspond to the common redundancy levels that one can find for data centers: N+1, 2N and 2(N+1) [49], and the configuration with N+2 is created to study opportunities for better reliability with more redundant equipment. There are also configurations with two chiller plants to simulate data centers with alternative chiller plants, like the configuration in Figure 3 [2]. When a chiller plant fails, the other one can be used to provide cooling continuously. Since each plant has a separate control box and each plant has the capability to support the cooling of the entire data center, the plants will not be operated simultaneously. The plants share the same water supply if both are water-cooled, and both fail if the water supply becomes unavailable.

Table 3 also considers the use of air-cooled chillers because the extra chiller plant in some Tier IV data centers uses air-cooled chillers as their extra equipment instead of water-cooled chillers. This avoids service interruption due to water deficiency [2]. They have no dependence on cooling water pumps and cooling towers because they reject waste heat to air directly at the chillers.

Table 3 also contains two configurations which 2N and 2(N+1) equipment are installed in one single plant with a redundant controller. They simulate the situation of a Tier IV data center which two chiller plants operate simultaneously as if the equipment belongs to one single chiller plant. Mathematically, the failure rate and the availability of the configurations can be described by Equations (25) and (26) respectively.

$$\lambda_{sys,simop} = \lambda_{par}(m = 1, M = 2, \lambda_{ctrl}, \mu_{ctrl}) - \lambda_{ctrl} + \lambda_{plant} + \lambda_{CRAH,gp}(2m_{ch}, 2M_{ch}) \quad (25)$$

$$p_{sys,nor,simop} = \frac{p_{nor,par}(m = 1, M = 2, p_{ctrl})}{p_{ctrl}} p_{plant,nor} p_{CRAH,gp,nor}(2m_{ch}, 2M_{ch}) \quad (26)$$

3.4 Configurations with distribution headers in a chiller plant

When distribution headers are installed in a chiller plant, some equipment are no longer dedicated to one chiller only. Even when one of them fails, the other equipment of the same type can be operated to avoid service interruption. Figure 5 illustrates how the distribution headers can be installed in the baseline scenario as shown in Figure 4. It shows three possible locations of extra distribution headers to enhance the reliability of a chiller plant: cooling tower distribution headers, cooling water pump distribution headers and chilled water pump distribution headers. The number of operating pumps and cooling towers can also be manipulated to be different from that of the number of operating chillers for higher energy efficiency [41], [48].

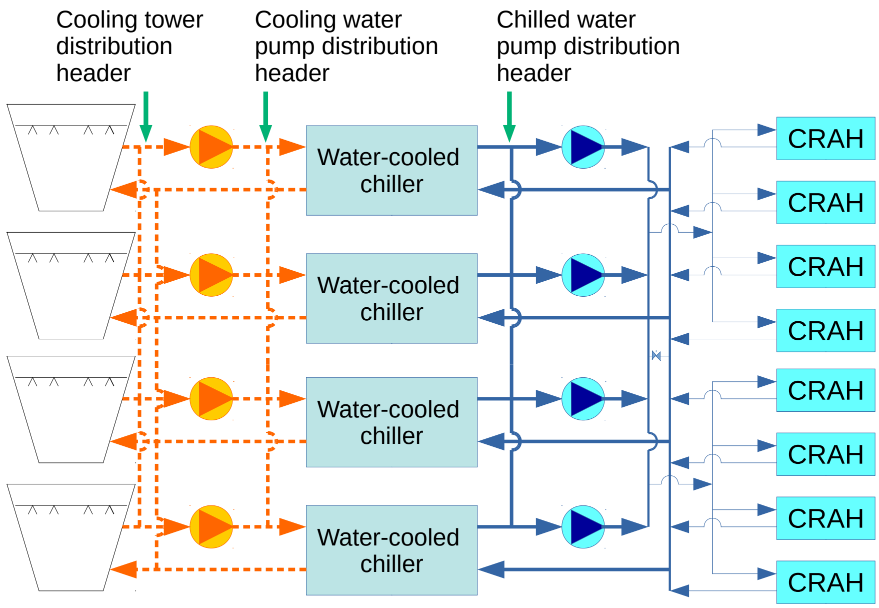

Figure 5: Possible locations of extra distribution headers

In this study, the baseline scenario in Figure 4 is configured to use the distribution headers shown in Figure 5 under conditions as Table 4. In the first configuration in Table 4, only the cooling tower distribution header shown in Figure 5 would be used in the cooling system without other distribution headers. In the second configuration, the cooling water circuit would contain the distribution headers for the cooling water pump and the cooling towers. The chilled water pump distribution header would only be added in the third configuration. The fourth and fifth configurations simulate the situation which the number of pumps required for operation is lower than the number of operating chillers for optimal control [41].

Table 4: Configurations under different installations and applications of distribution headers

Configuration with different installations and applications of distribution headers	Presence of cooling tower distribution header	Presence of cooling water pump distribution header	Presence of chilled water pump distribution header	Operating one fewer cooling water pump	Operating one fewer chilled water pump
1	Yes	No	No	No	No
2	Yes	Yes	No	No	No

3	Yes	Yes	Yes	No	No
4	Yes	Yes	No	Yes	No
5	Yes	Yes	Yes	Yes	Yes

3.5 Typical availability of individual equipment used in reliability assessment

The availabilities of individual equipment (e.g. chillers, pumps, cooling towers and CRAHs) can be calculated based on the mean time to failure and repair of their components from on-site observations as shown in Table 5.

Table 5 Mean time to failure (MTTF) and mean time to repair (MTTR) of components of chillers, pumps, cooling towers and CRAHs from on-site observations [50], [51]

Item	MTTF (hrs)	MTTR (hrs)	λ (per hour) based on Equation (2)	μ (per hour) based on Equation (3)
Chiller machinery	190891	14.52	5.239×10^{-6}	0.0689
Electric control valve	885743	18.42	1.129×10^{-6}	0.0543
Pump housing	1507745	6.75	6.632×10^{-7}	0.1481
Pump motor	6041379	3	1.655×10^{-7}	0.3333
Pump inverter	1817427	26	5.502×10^{-7}	0.0385
Pump variable-speed drive	396918	16.55	2.519×10^{-6}	0.0604
Cooling tower housing	1505154	16.67	6.644×10^{-7}	0.0600
Cooling tower fan blades	866468	7.91	1.154×10^{-6}	0.1264
Fan motor	791327	1	1.264×10^{-6}	1.0000
Fan variable-speed drive	396918	16.55	2.519×10^{-6}	0.0604
Switchgear control panel	446483	1.27	2.240×10^{-6}	0.7874
Piping (per unit)	833333	2	1.200×10^{-6}	0.5000
Air handling unit without motor drive	796363	2.36	1.256×10^{-6}	0.4237
Water utility	10862	12.68	9.206×10^{-5}	0.0789

To calculate the availability of the individual equipment, the number of components per equipment is needed. The information is tabulated in Table 6.

Table 6 List of components in different equipment

Equipment	Components
Water-cooled chiller	Chiller machinery x1, Electric control valve x2, Piping x 4, Switchgear control panel x1
Cooling tower	Cooling tower housing x1, Cooling tower fan blades x1, Fan motor x1, Fan variable-speed drive x1, Electric control valve x3, Piping x3, Switchgear control panel x1

CRAH	Air handling unit without motor drive x1, Fan motor x1, Electric control valve x1, Piping x2, Switchgear control panel x1
Variable-speed chilled water pump	Pump housing x1, Pump motor x1, Pump inverter x1, Pump variable-speed drive x1, Switchgear control panel x1, Electric control valve x1, Piping x2
Single-speed cooling water pump	Pump housing x1, Pump motor x1, Pump inverter x1, Switchgear control panel x1, Electric control valve x1, Piping x2
Air-cooled chiller	Chiller machinery x1, Electric control valve x2, Piping x2, Fan motor x1, Switchgear control panel x1

The list of equipment in Table 6 can be assumed to be connected in series in each equipment for the calculation of the failure rates, the availabilities and the reliabilities in a year for each equipment in Figure 4 based on Equations (1), (8) and (9). The results are listed in Table 7.

Table 7: Failure rates, reliabilities after a year of operation and availabilities of important components of a data center cooling system

Component	Failure rate (per hour)	Reliability after a year of operation	Availability
Air-cooled chiller	1.340×10^{-5}	88.9244%	99.9873%
Cooling tower	1.483×10^{-5}	87.8187%	99.9864%
CRAH	9.609×10^{-6}	91.9271%	99.9946%
Single-speed cooling water pump	7.148×10^{-6}	93.9306%	99.9952%
Switchgear control panel	2.240×10^{-6}	98.0571%	99.9997%
Variable-speed chilled water pump	9.667×10^{-6}	91.8803%	99.9911%
Water-cooled chiller	1.454×10^{-5}	88.0436%	99.9870%
Water utility	9.206×10^{-5}	44.6426%	99.8834%

3.6 Assessment procedure of the effect of different configurations to system reliability

To analyze the effects of various configurations and operating conditions to a data center cooling system, the configurations in Table 2, Table 3 and Table 4 are combined in various fashions to form multiple scenarios. The availabilities of the data center cooling system in these scenarios can be calculated using the availability of the individual equipment and Equations (9), (10), (13) and (14). The failure rates and reliabilities of the data center cooling system in these scenarios are calculated using Equations (1), (7), (8) and (12). The detailed calculation steps of the reliabilities and the availabilities of some complex scenarios are listed in the Appendix for reference.

4. Reliability and availability of different scenarios and enhancement of system reliability and availability by different configurations

By calculating the failure rates and the availabilities of the data center cooling systems in different scenarios and their reliabilities of the cooling systems after a year of operation from their failure rates by Equation (1), the effects of increasing cooling load, the effects of extra chillers, the effects of extra chiller plant, the effects of the use of both air-cooled and water-cooled systems, and the effects of adding distribution headers to a data center cooling system on the reliability and availability of a data center can be evaluated.

4.1 Effects of increasing cooling load

The effects of different cooling load on the data center cooling system reliability and availability can be studied by calculating the change of reliability after a year of operation and availability with the number of chillers required to operate in the baseline scenario in Figure 4. The results in Table 8 show that the reliability of the cooling system after a year of operation is above 43% and the availability of the cooling system is 99.88% in scenarios that do not need to run all chillers. When all chillers are needed, the reliability of the system after a year of operation falls to 4.426% and the availability of the system only falls to 99.68%. This shows that the cooling load of a data center has negligible effects on the system reliability and availability unless it needs to run all chillers in the system, but the cooling system is very likely to meet a failure and cannot meet the service requirement in a year if the operation all chillers are always required.

Table 8: Effect of the number of required chillers of system reliability and availability

Number of required chillers	Reliability after a year of operation	Availability
1 (N-3)	43.775%	99.883%
2 (N-2)	43.775%	99.883%
3 (N-1)	43.690%	99.883%
4 (N)	4.426%	99.679%

Table 8 also shows that the baseline scenario fails to meet the requirement of Tier II, III and IV data centers in Table 1. In Table 8, the minimum availability of the data center cooling system in the baseline scenario is 99.68%. This value is lower than the required availabilities of service in Tier II, III and IV data centers in Table 1. Since the availability of their data centers must be higher than the required overall availability of service to meet the requirement in the tier classification, the data center cooling system in the baseline scenario, which has N equipment, cannot be used in Tier II, III and IV data centers.

4.2 Effects of redundant chillers and associated equipment

The effect of a redundant chiller and associated equipment on the reliability and availability of a data center cooling system can be studied by calculating the reliability and availability of the data center cooling system with N+1 chillers and N+2 chillers instead of the N chillers in the baseline scenario. The calculation results in Table 9 shows similar results as Table 8 with the exception of the last scenario. In the last scenario, the cooling system requires N chillers to operate. The corresponding reliability in a year has increased from 4.4262% to a level above 43%, and the corresponding availability has increased from 99.679% to 99.883%. By having a redundant chiller, the reliability of the data center cooling system in a year remains above 43% at all cooling loads, and its availability also remains at 99.883% at all cooling loads. The system can satisfy the requirement of both Tier I and II data centers in Table 1.

Table 9 Reliabilities and availabilities of a data center cooling system under N (4 duty chillers), N+1 (4 duty chillers and 1 stand-by chiller) and N+2 configurations

Number of required chillers	Reliability after a year of operation			Availability		
	N equipment	N+1 equipment	N+2 equipment	N equipment	N+1 equipment	N+2 equipment
1 (N-3)	43.7753%	43.7753%	43.7753%	99.8831%	99.8831%	99.8831%
2 (N-2)	43.7753%	43.7753%	43.7753%	99.8831%	99.8831%	99.8831%
3 (N-1)	43.6897%	43.7752%	43.7753%	99.8831%	99.8831%	99.8831%
4 (N)	4.4262%	43.6328%	43.7751%	99.6792%	99.8830%	99.8831%

Table 9 also shows that the system reliability and availability continue to increase when the number of available equipment in a system increases from N+1 to N+2. But the gains of system reliability after a year of operation and system availability are only 0.1423% and 0.0001% respectively in the last scenario and are largely negligible.

4.3 Effects of an extra chiller plant

To study the effect of an extra chiller plant, the reliabilities and availabilities of the data center cooling system with N operating chillers but different number of duty and stand-by chillers and chiller plants are calculated. The results in Table 10 show that an extra chiller plant and the simultaneous operation of chiller plants can increase the reliability and availability of a data center cooling system beyond that of having an extra chiller. But the availability of the data center cooling system with a redundant water-cooled chiller plant remains smaller than the required availability of service of Tier III data centers in Table 1. An extra water-cooled chiller plant and simultaneous operation of both chiller plants are only capable to help a data center cooling system to meet the Tier II data center requirement but not the ones of Tiers III and IV data centers.

Table 10: Reliabilities and availabilities in scenarios with extra chillers and chiller plants

Total number of chillers	Presence of an extra chiller plant	Presence of an extra controller	Reliability after a year of operation	Availability
4 (N)	No	No	4.4262%	99.6792%
5 (N+1)	No	No	43.6328%	99.8830%
6 (N+2)	No	No	43.7751%	99.8831%
8 (2N)	Yes	No	44.2218%	99.8830%
10 (2(N+1))	Yes	No	44.6226%	99.8834%
8 (2N)	No	Yes	45.5271%	99.8837%
10 (2(N+1))	No	Yes	45.5271%	99.8837%

4.4 Effects of extra air-cooled chillers

The effect of using air-cooled chillers in the extra chiller plant on the reliability and availability of the data center cooling system can be seen in Table 11. Table 11 shows a significant increase of availability from 99.88% in cases without air-cooled chiller to 100.00% in cases with air-cooled chillers. The reliability of the cooling system after a year of operation is also doubled by an extra air-cooled chiller. The main reason is that the air-cooled chiller does not require water utility to operate, and the independence avoids the data center from total failures due to total failure in water supply. The extra air-cooled chillers also allow the data center cooling system to meet the requirement of Tier III and IV data centers in Table 1.

Table 11 Reliability and availability of data center cooling systems using air-cooled chillers in the extra chiller plants

Total number of chillers	Using air-cooled chiller in the extra chiller plant	Reliability after a year of operation	Availability
8 (2N)	No	44.2218%	99.8830%
8 (2N)	Yes	99.1154%	(100 – 0.000417) %
10 (2(N+1))	No	44.6426%	99.8834%
10 (2(N+1))	Yes	99.9973%	$(100 - 3.87 \times 10^{-7})$%

To examine if the water utility availability is the major cause of the low reliability and availability for water-cooled chiller plants, the values in Table 11 are re-evaluated by assuming that the sources of cooling water of the two water-cooled chiller plants are separated. The failure rates and the availabilities of components and water utility are still maintained at the values in Table 7, but the calculation of the availability of the data center cooling systems are no longer adjusted for the shared water source. The results in Table 12 shows that the use of a different source of cooling water increases the reliability and availability of the data center cooling system to the same level as that with air-cooled chillers in Table 11. Hence data center cooling systems in Tier III and IV data centers should use separate medium of waste heat rejection for their chiller plants.

Table 12 Reliability and availability of the data center cooling system by using separate cooling water sources for chiller plant heat rejection

Total number of chillers	Presence of an extra plant	Using separate water sources for the cooling water in the extra plant	Reliability after a year of operation	Availability
8 (2N)	Yes	Yes	98.0132%	(100 – 0.000592) %
10 (2(N+1))	Yes	Yes	99.8058%	$(100 – 1.55 \times 10^{-5})$%

4.5 Effects of distribution headers

Since distribution headers only increase the system availabilities in scenarios which the number of required chillers is fewer than the number of available chillers, the effect of distribution headers is studied with N+1 equipment and N operating chillers only. The results in Table 13 suggests that the increase of availability due to distribution headers is only around 0.00008% which is much smaller than the 0.2% increase of availability brought by an additional water-cooled chiller or an extra chiller plant. The increase of reliability by distribution headers is also less than 0.1% which is much smaller than the increase due to an extra water-cooled chiller or an extra chiller plant. The gains of reliability and availability by reducing the number of operating pumps through control are also similar and are negligible.

Table 13 Reliability and availability of the data center cooling system with N+1 equipment

Presence of cooling tower distribution header	Presence of cooling water pump distribution header	Presence of chilled water pump distribution header	Operating one fewer cooling water pump	Operating one fewer chilled water pump	Reliability after a year of operation	Availability
No	No	No	No	No	43.63279%	99.88295%
Yes	No	No	No	No	43.69565%	99.88303%
Yes	Yes	No	No	No	43.71651%	99.88305%
Yes	Yes	Yes	No	No	43.73610%	99.88307%
Yes	Yes	No	Yes	No	43.71912%	99.88305%
Yes	Yes	Yes	Yes	Yes	43.74534%	99.88308%

The reason for the small changes of availability due to distribution headers is the small number of cases which the distribution headers are the necessary equipment for the cooling system operation. When there are N+1 equipment, distribution headers are only necessary when more than two different types of equipment fail simultaneously. If there are more than two equipment of the same type fail, an extra chiller plant will be needed to sustain the normal operation. The number of situation that the distribution headers are necessary for the operation is small, and hence the distribution headers cannot enhance the reliability of a data center cooling system significantly.

5. Conclusions

A study on how configurations of water-cooled data center cooling systems can be improved to satisfy the reliability requirements of data centers of different tiers is conducted. The study is carried out by using the constant failure rate and the steady-state availability of data center cooling equipment from the field and calculating the availability of the entire systems with different numbers of redundant equipment and distribution headers.

The results show that a water-cooled chiller plant can satisfy the requirements of Tier II data centers by having one redundant chiller, and it can satisfy the requirements of Tier III and IV data centers by having one redundant chiller plant with a different source of cooling supply such as air or another source of cooling water. In particular, the use of a different source of cooling supply can double the reliability of a cooling system after a year of operation. The results also show that the availabilities of the N+2 (N duty + 2 stand-by) configuration and 2(N+1) configuration are only 0.0001% and 0.0003% higher than that of the N+1 configuration and the 2N configuration respectively and are not necessary to increase the availability of a data center cooling system. The results also show that each additional pair of distribution headers between pump, chillers and cooling towers in a water-cooled chiller plant increase the reliability and availability of a data center cooling system by 0.02% and 0.00008% only respectively. The increases are not as significant as the use of redundant equipment, and the use of additional distribution headers cannot replace redundant equipment as a major measure to enhance the availability of data center cooling systems.

While the study assesses how different redundant equipment and distribution header designs achieve the availability requirement of different tiers of data centers, the study can be improved by accounting for the effects of cooling water storages. Further research work can also be made to study the effects of various control algorithms on system reliability.

6. Acknowledgements

This research is financially supported by The Hong Kong Polytechnic University Postdoctoral Fellowship (G-YW2B) of The Hong Kong Polytechnic University and a grant under the Strategic Focus Area (SFA) Scheme of the Research Institute for Sustainable Urban Development (RISUD) in The Hong Kong Polytechnic University.

Nomenclature

Roman	
CRAC	Computer room air conditioner
CRAH	Computer room air handler
M	Total number of equipment in a chiller plant
m	Number of required operating equipment
MTTF	Mean time to failure [hour]
MTTR	Mean time to repair [hour]
N	Minimum number of equipment required in a chiller plant
n_{series}	Number of equipment connected in series
P	Probability matrix
p	Probability
t	Time [hour]
R	Reliability
Greek	
λ	Failure rate [/hour]
μ	Repair rate [/hour]
Subscript	
ach	Air-cooled chiller
aplant	Air-cooled chiller plant
cdd	Cooling water pump distribution header
cdp	Cooling water pump
ch	Chiller
chd	Chilled water pump distribution headers
chp	Chilled water pump
CRAH	Computer room air handler
ctd	Cooling tower distribution headers
ctrl	Control box
f	Failed state
gp	Group
ij	From state i to state j
nor	Normal state
only	Only
opt	Optimal
par	Parallel configuration
plant	Chiller plant

series	Series configuration
simop	Simultaneous operation
water	Water utility
wch	Water-cooled chiller
wplant	Water-cooled chiller plant
ww	Without water utility

References

[1] M. Stansberry, *Uptime Institute's 2017 Data Center Industry Survey Results*. Uptime Institute, 2017.

[2] W. P. Turner, J. H. Seader, V. Renaud, and K. G. Brill, "Tier classification define site infrastructure performance," Uptime Institute, 2006.

[3] W. Turner, J. H. Seader, and W. Renaud, "Data center site infrastructure tier standard: Topology," Uptime Institute, 2010.

[4] D. Humphrey, R. P. Domingo, M. A. Bemat, and X. Zhu, "Redundant power supply systems and methods," US Patent, 29-Jul-2009.

[5] J. Grimshaw, "Datacenter utilizing modular infrastructure systems and redundancy protection from failure," US Patent, 14-Jul-2011.

[6] S. B. Bekiarov and A. Emadi, "Uninterruptible power supplies: classification, operation, dynamics, and control," in *APEC. Seventeenth Annual IEEE Applied Power Electronics Conference and Exposition (Cat. No.02CH37335)*, 2002, vol. 1, pp. 597–604 vol.1.

[7] B. Hammond, C. J. Meiser, T. Giaquinto, C. Murphy, and D. Redmond, "System and method for managing uninterruptible power supply systems," US Patent, 22-Mar-2004.

[8] Q. Fang, J. Wang, and Q. Gong, "QoS-Driven Power Management of Data Centers via Model Predictive Control," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 4, pp. 1557–1566, Oct. 2016.

[9] T. Chen, X. Wang, and G. B. Giannakis, "Cooling-Aware Energy and Workload Management in Data Centers via Stochastic Optimization," *IEEE J. Sel. Top. Signal Process.*, vol. 10, no. 2, pp. 402–415, Mar. 2016.

[10] Y. Song, H. Wang, Y. Li, B. Feng, and Y. Sun, "Multi-Tiered On-Demand Resource Scheduling for VM-Based Data Center," in *2009 9th IEEE/ACM International Symposium on Cluster Computing and the Grid*, 2009, pp. 148–155.

[11] F. Machida, M. Kawato, and Y. Maeno, "Redundant virtual machine placement for fault-tolerant consolidated server clusters," in *2010 IEEE Network Operations and Management Symposium - NOMS 2010*, 2010, pp. 32–39.

[12] F. Favoino, A. Capozzoli, and M. Perino, "Temperature Field Real-Time Diagnosis by Means of Infrared Imaging in Data Elaboration Center," in *Proceedings of the 8th International Symposium on Heating, Ventilation and Air Conditioning*, Springer, Berlin, Heidelberg, 2014, pp. 455–463.

[13] M. Nazari Cheraghlou, A. Khadem-Zadeh, and M. Haghparast, "A survey of fault tolerance architecture in cloud computing," *J. Netw. Comput. Appl.*, vol. 61, pp. 81–92, Feb. 2016.

[14] M. Smara, M. Aliouat, A.-S. K. Pathan, and Z. Aliouat, "Acceptance Test for Fault Detection in Component-based Cloud Computing and Systems," *Future Gener. Comput. Syst.*, vol. 70, pp. 74–93, May 2017.

[15] M. Lin, S. Shao, X. (Simon) Zhang, J. W. VanGilder, V. Avelar, and X. Hu, "Strategies for data center temperature control during a cooling system outage," *Energy Build.*, vol. 73, pp. 146–152, Apr. 2014.

[16] M. Al-Fares, A. Loukissas, and A. Vahdat, "A Scalable, Commodity Data Center Network Architecture," in *Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication*, New York, NY, USA, 2008, pp. 63–74.

[17] R. Niranjan Mysore *et al.*, "PortLand: A Scalable Fault-tolerant Layer 2 Data Center Network Fabric," in *Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication*, New York, NY, USA, 2009, pp. 39–50.

[18] A. Greenberg *et al.*, "VL2: A Scalable and Flexible Data Center Network," in *Proceedings of the ACM SIGCOMM 2009 Conference on Data Communication*, New York, NY, USA, 2009, pp. 51–62.

[19] A. Putnam *et al.*, “A Reconfigurable Fabric for Accelerating Large-scale Datacenter Services,” in *Proceeding of the 41st Annual International Symposium on Computer Architecuture*, Piscataway, NJ, USA, 2014, pp. 13–24.
[20] H. Xu, C. Feng, and B. Li, “Temperature Aware Workload Managementin Geo-Distributed Data Centers,” *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 6, pp. 1743–1753, Jun. 2015.
[21] T. Chen, Y. Zhang, X. Wang, and G. B. Giannakis, “Robust geographical load balancing for sustainable data centers,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 3526–3530.
[22] R. Tripathi, V. S, V. Tamarapalli, and D. Medhi, “Cost Efficient Design of Fault Tolerant Geo-Distributed Data Centers,” *IEEE Trans. Netw. Serv. Manag.*, vol. PP, no. 99, pp. 1–1, 2017.
[23] R. Menuet and W. P. Turner, “Continuous cooling is required for continuous availability,” Uptime Institute, 2006.
[24] N. Rasmussen, “Guidelines for Specification of Data Center Power Density,” American Power Conversion, 2005.
[25] Z. Hasan, “Redundancy for Data Centers,” *ASHRAE J.*, vol. 51, no. 2, pp. 52–54, Feb. 2009.
[26] X. Zhang, L. Jia, J. Wu, R. Wang, J. Li, and Y. Zhong, “Efficient Water-Cooled Chillers,” in *Handbook of Energy Systems in Green Buildings*, R. Wang and X. Zhai, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2018, pp. 755–798.
[27] LBNL, “Data Center Master List of Energy Efficiency Actions | Center of Expertise for Energy Efficiency in Data Centers,” Lawrence Berkerly National Laboratory, 2016.
[28] F. D. Angelis and U. Grasselli, “A Data Center performance comparison analysis between Tier III architecture and a CCHP alternative solutions,” in *2015 5th International Youth Conference on Energy (IYCE)*, 2015, pp. 1–6.
[29] D. Xu and M. Qu, “Energy, environmental, and economic evaluation of a CCHP system for a data center based on operational data,” *Energy Build.*, vol. 67, no. Supplement C, pp. 176–186, Dec. 2013.
[30] M. Wiboonrat, “An Empirical Study on Data Center System Failure Diagnosis,” in *2008 The Third International Conference on Internet Monitoring and Protection*, 2008, pp. 103–108.
[31] J. Stein, “Waterside Economizing in Data Centers: Design and Control Considerations,” *ASHRAE Trans.*, vol. 115, pp. 192–200, 2009.
[32] R. Zhou, Z. Wang, C. E. Bash, T. Cader, and A. McReynolds, “Failure Resistant Data Center Cooling Control Through Model-Based Thermal Zone Mapping,” in *ASME 2012 Heat Transfer Summer Conference*, 2012, pp. 751–757.
[33] J. Dai, D. Das, M. Ohadi, and M. Pecht, “Reliability risk mitigation of free air cooling through prognostics and health management,” *Appl. Energy*, vol. 111, pp. 104–112, Nov. 2013.
[34] H. A. Alissa *et al.*, “Chip to Chiller Experimental Cooling Failure Analysis of Data Centers: The Interaction Between IT and Facility,” *IEEE Trans. Compon. Packag. Manuf. Technol.*, vol. 6, no. 9, pp. 1361–1378, Sep. 2016.
[35] M. Iyengar, R. Schmidt, and J. Caricari, “Reducing energy usage in data centers through control of Room Air Conditioning units,” in *2010 12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, 2010, pp. 1–11.
[36] S. A. Bermudez, H. F. Hamann, L. J. Klein, F. J. Marianno, and A. Claassen, “Optimal and distributed automatic discrete control of air conditioning units in data centers,” in *2015 31st Thermal Measurement, Modeling Management Symposium (SEMI-THERM)*, 2015, pp. 13–18.
[37] J.-J. Wang, C. Fu, K. Yang, X.-T. Zhang, G. Shi, and J. Zhai, “Reliability and availability analysis of redundant BCHP (building cooling, heating and power) system,” *Energy*, vol. 61, pp. 531–540, Nov. 2013.
[38] G. Abdollahi and M. Meratizaman, “Multi-objective approach in thermoenvironomic optimization of a small-scale distributed CCHP system with risk analysis,” *Energy Build.*, vol. 43, no. 11, pp. 3144–3153, Nov. 2011.
[39] W. Gang, S. Wang, F. Xiao, and D. Gao, “Robust optimal design of building cooling systems considering cooling load uncertainty and equipment reliability,” *Appl. Energy*, vol. 159, pp. 265–275, Dec. 2015.

[40] M. Wiboonrat, “Zone protection approach of data center system reliability,” in *2017 International Conference on Information and Communications (ICIC)*, 2017, pp. 240–246.

[41] J. E. Braun, “Methodologies for the design and control of central cooling plants,” PhD Thesis, University of Wisconsin-Madison, 1988.

[42] Z. Liu, H. Tan, D. Luo, G. Yu, J. Li, and Z. Li, “Optimal chiller sequencing control in an office building considering the variation of chiller maximum cooling capacity,” *Energy Build.*, vol. 140, pp. 430–442, Apr. 2017.

[43] Y. Wang, X. Jin, and X. Fang, “Rapid evaluation of operation performance of multi-chiller system based on history data analysis,” *Energy Build.*, 2016.

[44] C. E. Ebeling, *An introduction to reliability and maintainability engineering*. New York: McGraw Hill, 1997.

[45] Q. Cheng, S. Wang, C. Yan, and F. Xiao, “Probabilistic approach for uncertainty-based optimal design of chiller plants in buildings,” *Appl. Energy*, vol. 185, no. Part 2, pp. 1613–1624, Jan. 2017.

[46] A. Lisnianski, I. Frenkel, and Y. Ding, *Multi-state System Reliability Analysis and Optimization for Engineers and Industrial Managers*. London: Springer London, 2010.

[47] R. E. Walpole, R. H. Myers, S. L. Myers, and K. E. Ye, *Probability and Statistics for Engineers and Scientists*, 9th edition. Boston: Pearson, 2011.

[48] ASHRAE, *2016 ASHRAE Handbook HVAC Systems and Equipment*. Atlanta, GA: American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc., 2016.

[49] S. W. Duda, “N+1 HVAC for IT closets and server rooms,” *ASHRAE J.*, vol. 60, no. 5, pp. 56–60, 2018.

[50] K. W. Mui, L. T. Wong, and K. W. Hui, “Downtime of in-use water pump installations for high-rise residential buildings,” *Build. Serv. Eng. Res. Technol.*, vol. 33, no. 2, pp. 181–190, May 2012.

[51] P. S. Hale and R. G. Arno, “Survey of reliability and availability information for power distribution, power generation, and HVAC components for commercial, industrial, and utility installations,” in *2000 IEEE Industrial and Commercial Power Systems Technical Conference. Conference Record (Cat. No.00CH37053)*, 2000, pp. 31–54.

[52] Emerson Network Power, “Understanding the Cost of Data Center Downtime: An Analysis of the Financial Impact on Infrastructure Vulnerability,” A White Paper from the Experts in Business-Critical Continuity, 2011.

[53] A. Birolini, *Reliability Engineering*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014.

[54] J. Wang, Q. Zhang, S. Yoon, and Y. Yu, “Reliability and availability analysis of a hybrid cooling system with water-side economizer in data center,” *Building and Environment*, vol. 148, pp. 405–416, Jan. 2019.

[55] R. Nishida, S. Waragai, K. Sekiguchi, M. Kishita, H. Miyake, and T. Uekusa, “Relationship between the reliability of a Data-center air-conditioning system and the air-conditioning power supply,” in INTELEC 2008 - 2008 IEEE 30th International Telecommunications Energy Conference, San Diego, CA, 2008, pp. 1–5.

[56] A. Lawrence, “Uptime Institute data shows outages are common, costly, and preventable,” Uptime Institute, 2018.

[57] T. A. Nguyen, D. Min, E. Choi, and T. D. Tran, “Reliability and Availability Evaluation for Cloud Data Center Networks Using Hierarchical Models,” IEEE Access, vol. 7, pp. 9273–9313, 2019.

[58] D. M. Gomes et al., “Evaluating the cooling subsystem availability on a Cloud data center,” in 2017 IEEE Symposium on Computers and Communications (ISCC), 2017, pp. 736–741.

[59] G. Callou, P. Maciel, D. Tutsch, and J. Araújo, “Models for dependability and sustainability analysis of data center cooling architectures,” in IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN 2012), 2012, pp. 1–6.

Appendix Calculation method of failure rates and availabilities for different settings

This appendix describes the calculation method of the failure rate and the availability of a chiller plant changes when air-cooled chillers are used, the calculation method of the failure rate and the availability of a chiller plant in scenarios with different distribution headers in Table 4, and the calculation method of the failure rate and the availability of a cooling system when it has more than one chiller plant as the cases in Table 3.

Scenarios with air-cooled chillers instead of water-cooled chillers in a chiller plant

In this study, when air-cooled chillers are used instead of water-cooled chillers in a chiller plant, all its chillers will be replaced by air-cooled chillers. The configuration of the chiller plant will be changed to that in Figure 6.

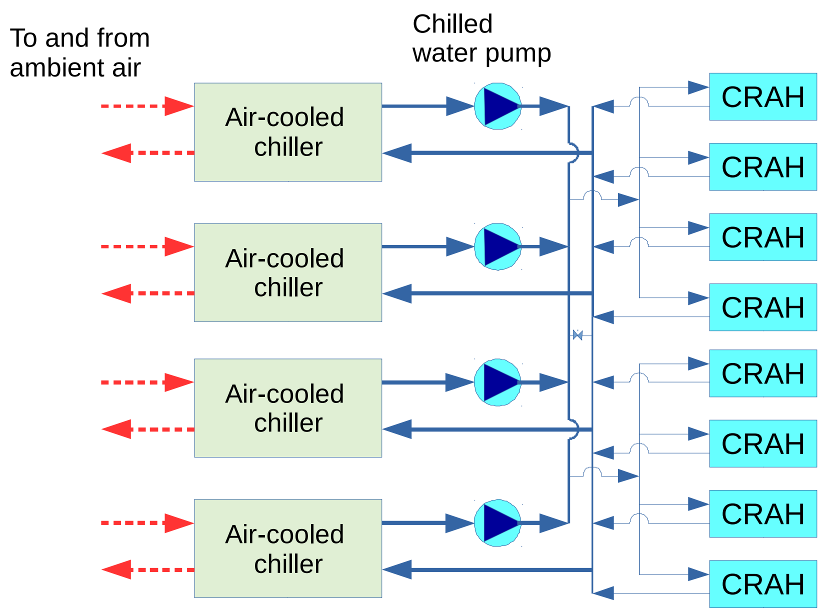

Figure 6 A cooling system with an air-cooled chiller plant and CRAHs

Figure 6 shows a chiller plant without cooling water pumps and cooling towers and the chiller is cooled by ambient air. Since the plant contains no cooling tower and cooling water pumps, the failure rate of the plant no longer depends on the availabilities of cooling towers and cooling water pumps from Equation (22). Only the failure rates of CRAHs, chilled water pumps and air-cooled chillers is needed to calculate the failure rate of an air-cooled chiller plant only as shown in Equations (27) and (29).

$$\lambda_{aplant,base}(m_{ch}, M_{ch}) = \lambda_{ctrl} + \lambda_{par}\left(m_{ch}, M_{ch}, \lambda_{ach} + \lambda_{chp}, \mu_{ach+chp}\right) \quad (27)$$

$$\mu_{ach+chp} = \mu\left(\lambda_{ach} + \lambda_{chp}, p_{ach,nor} p_{chp,nor}\right) \quad (28)$$

Equation (27) not only excludes the failure rates of cooling water pumps and cooling towers, it also excludes the failure rate of the water supply. The plant is no longer consuming water significantly through evaporative cooling, and it does not require a reliable water supply to operate. Similarly, the availability of an air-cooled chiller plant can be calculated by Equation (29).

$$p_{aplant,base,nor}(m_{ch}, M_{ch}) = p_{ctrl} p_{nor,par}\left(m_{ch}, M_{ch}, p_{ach,nor} p_{chp,nor}\right) \quad (29)$$

The number of distribution header scenarios that can be simulated with an air-cooled chiller plant are also reduced from that of water-cooled chiller plants because of the absence of the cooling water pumps and cooling towers. Since these two components are absent, only the distribution scenario III in Table 4 can be considered with the distribution header at the chilled water pumps only. To describe this scenario mathematically, the group of air-cooled chillers can be considered as having a series connection with the group of chilled water pumps, and Equations (8) and (12) can be used to derive Equation (30) for the failure rate of the air-cooled chiller plant with a distribution header at the chilled water pumps.

$$\lambda_{aplant,chd}(m_{ch}, M_{ch}) = \lambda_{par}(m_{ch}, M_{ch}, \lambda_{ach}, \mu_{ach}) + \lambda_{par}\left(m_{ch}, M_{ch}, \lambda_{chp}, \mu_{chp}\right) \quad (30)$$

Similarly, Equation (9) can be used to derive Equation (31) for the availability of the air-cooled chiller plant.

$$p_{aplant,chd,nor}(m_{ch}, M_{ch}) = p_{nor,par}\left(m_{ch}, M_{ch}, p_{ach,nor}\right) p_{nor,par}\left(m_{ch}, M_{ch}, p_{chp,nor}\right) \quad (31)$$

Scenarios with distribution headers around cooling towers only

To consider the failure rate and the availability of a chiller plant with distribution headers around its cooling towers, the group of cooling towers can be considered as being connected to the group of the rest of the equipment in series connection. The inter-equipment connection in the latter group remains the same. By using Equations (8) and (12), one can derive the failure rate equations of the chiller plant as Equations (32) and (33).

$$\begin{aligned}\lambda_{wplant,ctd}(m_{ch}, M_{ch}) \\ &= \lambda_{water} + \lambda_{ctrl} + \lambda_{par}\left(m_{ch}, M_{ch}, \lambda_{cdp} + \lambda_{wch} + \lambda_{chp}, \mu_{cdp+wch+chp}\right) \\ &+ \lambda_{par}(m_{ch}, M_{ch}, \lambda_{ct}, \mu_{ct})\end{aligned} \quad (32)$$

$$\mu_{cdp+wch+chp} = \mu\left(\lambda_{cdp} + \lambda_{wch} + \lambda_{chp}, p_{cdp,nor}p_{wch,nor}p_{chp,nor}\right) \quad (33)$$

By using Equations (9) and (10), one can derive the availability equation of the chiller plant as Equation (34).

$$\begin{aligned}p_{wplant,ctd,nor}(m_{ch}, M_{ch}) \\ &= p_{water}p_{ctrl}p_{nor,par}\left(m_{ch}, M_{ch}, p_{cdp,nor}p_{wch,nor}p_{chp,nor}\right) \\ &\times p_{nor,par}\left(m_{ch}, M_{ch}, p_{ct,nor}\right)\end{aligned} \quad (34)$$

Scenarios with distribution headers around cooling towers and cooling water pumps

In these scenarios, the group of cooling water pumps are considered to be in series connection with a group of cooling towers and the group of the rest of the equipment in the plant. The failure rate of the plant becomes Equations (35) and (36).

$$\begin{aligned}\lambda_{wplant,ctd+cdd}(m_{ch}, M_{ch}) \\ &= \lambda_{water} + \lambda_{ctrl} + \lambda_{par}\left(m_{ch}, M_{ch}, \lambda_{wch} + \lambda_{chp}, \mu_{wch+chp}\right) \\ &+ \lambda_{par}(m_{ch}, M_{ch}, \lambda_{ct}, \mu_{ct}) + \lambda_{par}\left(m_{ch}, M_{ch}, \lambda_{cdp}, \mu_{cdp}\right)\end{aligned} \quad (35)$$

$$\mu_{wch+chp} = \mu\left(\lambda_{wch} + \lambda_{chp}, p_{wch,nor}p_{chp,nor}\right) \quad (36)$$

The availability of the plant becomes Equation (37).

$$\begin{aligned}p_{wplant,ctd+cdd,nor}(m_{ch}, M_{ch}) \\ &= p_{water}p_{ctrl}p_{nor,par}\left(m_{ch}, M_{ch}, p_{wch,nor}p_{chp,nor}\right) \\ &\times p_{nor,par}\left(m_{ch}, M_{ch}, p_{ct,nor}\right)p_{nor,par}\left(m_{ch}, M_{ch}, p_{cdp,nor}\right)\end{aligned} \quad (37)$$

In the case which the number of operating cooling water pumps is one fewer than the number of operating chillers for optimal control following Braun [41], the equations become Equations (38) and (39).

$$\lambda_{wplant,ctd+cdd}(m_{ch}, M_{ch}) = \lambda_{water} + \lambda_{ctrl} + \lambda_{par}(m_{ch}, M_{ch}, \lambda_{wch,nor} + \lambda_{chp,nor}, \mu_{wch+chp}) + \lambda_{par}(m_{ch}, M_{ch}, \lambda_{ct}, \mu_{ct}) + \lambda_{nor,par}(m_{ch} - 1, M_{ch}, \lambda_{cdp}, \mu_{cdp}) \tag{38}$$

$$p_{wplant,ctd+cdd,nor}(m_{ch}, M_{ch}) = p_{water} p_{ctrl} p_{nor,par}(m_{ch}, M_{ch}, p_{wch,nor} p_{chp,nor}) \times p_{nor,par}(m_{ch}, M_{ch}, p_{ct,nor}) p_{nor,par}(m_{ch} - 1, M_{ch}, p_{cdp,nor}) \tag{39}$$

Scenarios with distribution headers around cooling towers, cooling water pumps and chilled water pumps

In these scenarios, the plant can be considered to have a series connection between a group of chilled water pumps, a group of chillers, a group of cooling water pumps and a group of cooling towers. The failure rate of the plant thus becomes Equation (40).

$$\lambda_{wplant,ctd+cdd+chd}(m_{ch}, M_{ch}) = \lambda_{water} + \lambda_{ctrl} + \lambda_{par}(m_{ch}, M_{ch}, \lambda_{chp}, \mu_{chp}) + \lambda_{par}(m_{ch}, M_{ch}, \lambda_{ch}, \mu_{ch}) + \lambda_{nor,par}(m_{ch}, M_{ch}, \lambda_{ct}, \mu_{ct}) + \lambda_{nor,par}(m_{op,ch}, M, \lambda_{cdp}, \mu_{cdp}) \tag{40}$$

The availability of the plant thus becomes Equation (41).

$$p_{wplant,ctd+cdd+chd,nor}(m_{ch}, M_{ch}) = p_{water} p_{ctrl} p_{nor,par}(m_{ch}, M_{ch}, p_{chp,nor}) p_{nor,par}(m_{ch}, M_{ch}, p_{ch,nor}) \times p_{nor,par}(m_{ch}, M_{ch}, p_{ct,nor}) p_{nor,par}(m_{op,ch}, M, p_{cdp,nor}) \tag{41}$$

In the case which the water-cooled chiller plant control is optimized by the algorithm in Braun [41], the number of operating pumps is one fewer than the number of operating chillers, and the equations become Equations (42) and (43).

$$\begin{aligned}\lambda_{wplant,ctd+cdd+chd,opt}(m_{ch}, M_{ch}) \\ = \lambda_{water} + \lambda_{ctrl} + \lambda_{par}(m_{ch} - 1, M_{ch}, \lambda_{chp}, \mu_{chp}) \\ + \lambda_{par}(m_{ch}, M_{ch}, \lambda_{ch}, \mu_{ch}) + \lambda_{par}(m_{ch}, M_{ch}, \lambda_{ct}, \mu_{ct}) \\ + \lambda_{par}(m_{ch} - 1, M, \lambda_{cdp}, \mu_{cdp})\end{aligned} \tag{42}$$

$$\begin{aligned}p_{wplant,ctd+cdd+chd,opt,nor}(m_{ch}, M_{ch}) \\ = p_{water} p_{ctrl} p_{nor,par}(m_{ch} - 1, M_{ch}, p_{chp,nor}) p_{nor,par}(m_{ch}, M_{ch}, p_{ch,nor}) \\ \times p_{nor,par}(m_{ch}, M_{ch}, p_{ct,nor}) p_{nor,par}(m_{ch} - 1, M, p_{cdp,nor})\end{aligned} \tag{43}$$

For air-cooled chiller plants, the equations become Equations (44) and (45).

$$\begin{aligned}\lambda_{aplant,ctd+cdd+chd}(m_{ch}, M_{ch}) \\ = \lambda_{ctrl} + \lambda_{par}(m_{ch}, M_{ch}, \lambda_{chp}, \mu_{chp}) + \lambda_{par}(m_{ch}, M_{ch}, \lambda_{ach}, \mu_{ach})\end{aligned} \tag{44}$$

$$\begin{aligned}p_{aplant,ctd+cdd+chd,nor}(m_{ch}, M_{ch}) \\ = p_{ctrl} p_{nor,par}(m_{ch}, M_{ch}, p_{chp,nor}) p_{nor,par}(m_{ch}, M_{ch}, p_{ach,nor})\end{aligned} \tag{45}$$

Scenarios with 2 chiller plants

In these scenarios, considering that only one plant is needed for all operation and the two chiller plants may be different, the calculation of the failure rate of a data center cooling system with two different plants will be processed by Equations (46), (47) and (48) [54].

$$\begin{aligned}\lambda_{sys}|_{2differentplants,m_{plant}=1}(m_{ch}, M_{ch}) \\ = \left(\sum_{i=1}^{2} \frac{1}{\mu_{plant,i+CRAH,gp}(m_{ch}, M_{ch})}\right) \prod_{i=1}^{2} \lambda_{plant,i+CRAH,gp}(m_{ch}, M_{ch})\end{aligned} \tag{46}$$

$$\lambda_{plant,i+CRAH,gp}(m_{ch}, M_{ch}) = \lambda_{plant,i}(m_{ch}, M_{ch}) + \lambda_{CRAH,gp}(2m_{ch}, 2M_{ch}) \tag{47}$$

$$\begin{aligned}\mu_{plant,1+CRAH,gp}(m_{ch}, M_{ch}) \\ = \mu\left(\lambda_{plant,i+CRAH,gp}(m_{ch}, M_{ch}), p_{plant,i,nor}(m_{ch}, M_{ch}) p_{CRAH,gp,nor}(2m_{ch}, 2M_{ch})\right)\end{aligned} \tag{48}$$

The calculation of the availability of a data center cooling system with two different plants will be processed by Equation (49).

$$\begin{aligned}p_{sys,nor}|_{2differentplants,m_{plant}=1}(m_{ch}, M_{ch}) \\ = p_{plant,1,nor} p_{plant,2,nor} p_{CRAH,gp,nor}^{2} \\ + (1 - p_{plant,1,nor} p_{CRAH,gp,nor}) p_{plant,2,nor} p_{CRAH,gp,nor} \\ + p_{plant,1,nor} p_{CRAH,gp,nor}(1 - p_{plant,2,nor} p_{CRAH,gp,nor})\end{aligned} \tag{49}$$

In cases which all plants are water-cooled and share the same water supply, the equations are modified to describe the shared water supply. This is done by using Equations (51), (52) and (52).

$$\lambda_{sys}|_{2wplants,m_{plant}=1}(m_{ch}, M_{ch}) = \lambda_{water} + \left(\frac{2}{\mu_{(wplant+CRAH,gp),ww}(m_{ch}, M_{ch})}\right)\left(\lambda_{wplant+CRAH,gp}(m_{ch}, M_{ch}) - \lambda_{water}\right)^2 \tag{50}$$

$$\mu_{(wplant+CRAH,gp),ww}(m_{ch}, M_{ch}) = \mu\left(\lambda_{plant,i+CRAH,gp}(m_{ch}, M_{ch}) - \lambda_{water}, \frac{p_{plant,i,nor}(m_{ch}, M_{ch})p_{CRAH,gp,nor}(2m_{ch}, 2M_{ch})}{p_{water}}\right) \tag{51}$$

$$p_{sys,nor}|_{2wplants,m_{plant}=1}(m_{ch}, M_{ch}) = \frac{p_{wplant,nor}^2 p_{CRAH,gp,nor}^2}{p_{water}} + 2\left(1 - \frac{p_{wplant,nor}p_{CRAH,gp,nor}}{p_{water}}\right)p_{wplant,nor}p_{CRAH,gp,nor} \tag{52}$$