CoolProvision: Underprovisioning Datacenter Cooling

Ioannis Manousakis[‡] Íñigo Goiri[†*] Sriram Sankar[δ]
Thu D. Nguyen[‡] Ricardo Bianchini[†‡]

[‡]Rutgers University {im159, tdnguyen}@cs.rutgers.edu
[δ]GoDaddy sriram@godaddy.com
[†]Microsoft Research {inigog, ricardob}@microsoft.com

Abstract

Cloud providers have made significant strides in reducing the cooling capital and operational costs of their datacenters, for example, by leveraging outside air ("free") cooling where possible. Despite these advances, cooling costs still represent a significant expense mainly because cloud providers typically provision their cooling infrastructure for the worst-case scenario (*i.e.*, very high load and outside temperature at the same time). Thus, in this paper, we propose to reduce cooling costs by underprovisioning the cooling infrastructure. When the cooling is underprovisioned, there might be (rare) periods when the cooling infrastructure cannot cool down the IT equipment enough. During these periods, we can either (1) reduce the processing capacity and potentially degrade the quality of service, or (2) let the IT equipment temperature increase in exchange for a controlled degradation in reliability. To determine the ideal amount of underprovisioning, we introduce CoolProvision, an optimization and simulation framework for selecting the cheapest provisioning within performance constraints defined by the provider. CoolProvision leverages an abstract trace of the expected workload, as well as cooling, performance, power, reliability, and cost models to explore the space of potential provisionings. Using data from a real small free-cooled datacenter, our results suggest that CoolProvision can reduce the cost of cooling by up to 55%. We extrapolate our experience and results to larger cloud datacenters as well.

Categories and Subject Descriptors C.m [*Computer Systems Organizations*]: Miscellaneous; D.4.m [*Operating systems*]: Miscellaneous

Keywords cooling, datacenters, provisioning

* This work was done while Íñigo Goiri was at Rutgers University.

SoCC '15, August 27–29, 2015, Kohala Coast, HI, USA.
 ISBN 978-1-4503-3651-2/15/08...$15.00
DOI: http://dx.doi.org/10.1145/2806777.2806938

1. Introduction

Cloud datacenters are expensive to build and operate. The construction cost for large datacenters is roughly $9-13 per Watt of critical power, and is higher for small datacenters [5]. This may translate to over $100M for a 10MW datacenter. In addition, US datacenters collectively incur billions of dollars in electricity cost per year [40]. These high costs provide strong incentives for careful provisioning of datacenters and for improving their energy efficiency.

Cloud providers, such as Microsoft and Google, have made significant progress in improving their datacenters' energy efficiency to reduce operational costs (OPEX). For example, Google is reporting average Power Usage Effectiveness (PUE) values close to 1.1 in ideal locations [20], showing that it has reduced cooling OPEX significantly. Providers have achieved these improvements by combining multiple techniques, such as increasing the operating temperature of the IT equipment, completely isolating the cold and hot aisles, and leveraging outside air to cool the IT equipment. The latter technique is commonly called "free" cooling.

Reducing capital costs (CAPEX) has also received attention. For example, Fan *et al.* studied power provisioning, arguing that datacenters' current power delivery infrastructure is typically over-provisioned [17]. Govindan *et al.* studied the benefits of power underprovisioning when the datacenter uses batteries during peaks of power consumption [21].

However, to the best of our knowledge, no prior paper has explored the provisioning of the cooling equipment of modern datacenters, even though cooling CAPEX is still significant. Barroso *et al.* estimate that power and cooling account for approximately 80% of total construction cost [5]. Cooling CAPEX is particularly high in locations where free cooling must be combined with conventional (chiller-based) cooling for when the outside temperature is too high.

CAPEX costs are exacerbated by the fact that cloud providers typically provision cooling systems for the worst-case scenario, *i.e.* for periods of simultaneous high load and outside temperature. This may significantly increase the cost of cooling, while leaving much of the cooling capacity unused most of the time. Fortunately, it may be possible to lower costs by *underprovisioning* the cooling system, if the provider can tolerate some performance degradation in extreme circumstances and/or trade off some hardware reliability. For example, during periods of high outside temperature, some of the current workload may be deferred, consolidated into fewer machines, and/or the datacenter allowed to operate at higher than ideal temperatures. Such tradeoffs may lower both cooling CAPEX and OPEX, but may also incur extra costs with increased hardware replacements.

Thus, we propose CoolProvision, an approach for carefully underprovisioning the cooling system of a datacenter. The goal is to minimize the overall cost, including CAPEX, OPEX, and any reliability costs, within performance constraints defined by the cloud provider. CoolProvision includes: (1) cost models for provisioning and operating the cooling system, and hardware replacement; (2) cooling and reliability models representing the thermal behaviors and their impact on hardware failures; (3) performance and power models representing workload scheduling and energy management policies; and (4) an optimization and simulation framework that integrates the models and policies with weather data for the location, and the expected workload to compute the appropriate provisioning of the cooling system.

CoolProvision can lower costs through the use of cheaper cooling technologies (*e.g.*, free cooling even in hot and humid locations), as well as reducing the maximum cooling capacity required. Significant CAPEX savings also arise from the reduced peak power required by the datacenter's power delivery infrastructure (*e.g.*, UPS, generators, distribution capacity), but we do not address these savings in this paper.

As a real case study, we apply CoolProvision to the Parasol datacenter prototype [18]. Parasol comprises a small container that isolates cold and hot aisles, and combines a free cooling unit with a direct-expansion air conditioner (AC). Parasol has already been built so our study considers what the infrastructure provisioning could have been for it. Our results show that Parasol's cooling could have been underprovisioned for up to 55% cooling-related cost savings.

Though Parasol is obviously not a cloud datacenter, it does provide us the ability to create accurate thermal and cooling models, which we do not have for a real cloud datacenter, and experiment with free cooling and the workload and energy management techniques we consider. Nevertheless, we explore how CoolProvision and its models can be used or extrapolated for cloud datacenters as well.

2. Datacenter Cooling

In this section, we describe the main cooling approaches used in cloud datacenters: (1) chiller-based cooling, (2) water-side economized cooling, and (3) direct evaporative cooling (or simply free cooling). Table 1 summarizes the main characteristics of the approaches. At the end of the section, we also discuss the approaches used in small- and medium-scale datacenters, which are similar.

Chiller-based cooling. In this setup, datacenters use water chillers, cooling towers, and computer room air handlers (CRAHs). The CRAHs circulate the air carrying heat from the servers to cooling coils carrying chilled water. The heat is then transferred via the water back to the chillers, which transfer some of the heat to another water loop directed to the cooling towers. The towers cool the water down via evaporation. The chillers then remove the remaining heat from the water, and the chilled water is returned to the cooling coils. Chillers are expensive and consume a significant amount of energy. However, they allow precise control of the environmental conditions inside the datacenter (except for hot spots that may develop due to poor air flow design).

Technology	Temp./Hum. Control	CAPEX	PUE
Chiller	Precise / Precise	\$2.0/W	1.7
Water-side	Precise / Precise	\$2.8/W	1.19
Direct evap./free	Medium / Low	\$0.7/W	1.12

Table 1. Typical temperature and RH control, CAPEX costs [13, 31], and PUEs of the cooling technologies [16, 39].

Water-side economized cooling. An improvement over chiller-based cooling allows the chillers to be bypassed (and turned off) when the cooling towers alone are sufficient to cool the water. Turning the chillers off significantly reduces energy consumption. When the cooling towers cannot lower the temperature enough, the chillers come back on. Water-size economized cooling preserves the ability to precisely control the environmental conditions inside the datacenter. It also does not mix outside and inside air.

Direct evaporative (free) cooling. A more recent advance has been to use large fans to blow cool outside air into the datacenter, and guide the warm return air back out of it. When the outside temperature is high, an evaporative process can be applied [15], where water is sprayed on the outside air while it passes through a washer system. This process lowers the air temperature before letting it reach the servers. To increase temperature (during excessively cold periods) and/or reduce relative humidity, these datacenters intentionally recirculate some of the warm return air. Free cooling obviates the need for chillers and cooling towers, and thus is the least expensive to build and operate. However, it may also expose servers to warmer and more variable temperatures, and higher and more variable relative humidities than the other cooling approaches.

Small- and medium-scale datacenters. Cooling approaches for smaller datacenters are similar, although their free cooling systems may [14] or may not [18] include evaporative cooling. In addition, technologies such as direct-expansion air conditioning may be used instead of chillers and cooling towers. In fact, some smaller datacenters use a combination of free and direct-expansion cooling. Using such a combination, Parasol achieves a PUE of 1.13, which is comparable to that of free-cooled cloud datacenters.

3. Provisioning Cloud Datacenter Cooling

Traditional provisioning. Traditionally, the provisioning of a datacenter's cooling system starts with determining the

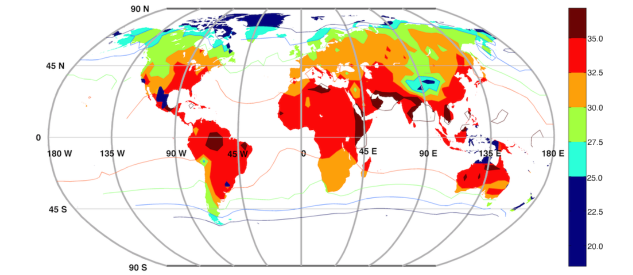

Figure 1. *Inlet air* temperature for a direct evaporative cooler. Red and brown areas show infeasible locations for free-cooled datacenters under ASHRAE's guidelines and traditional provisioning.

maximum desired cold aisle server inlet temperature and relative humidity, taking into consideration the servers' thermal limits, the reliability effects on server components, and the desired temperature differential between the cold and hot aisles. For example, the current ASHRAE guidelines allow inlet temperatures in the range of $15-32°C$ and relative humidity in the range of $20-80\%$ [23]. Next, the cooling approach (chiller-based vs. free cooling vs. hybrid) is selected based on a study of the historical weather data for the selected datacenter site. This process ensures that the cooling approach will *always* be able to meet the inlet temperature and humidity design requirements. Finally, the cooling capacity is determined, ensuring at least a 1:1 relationship to the *peak IT power consumption under the worst-case operating scenario*. For example, a 10 MW datacenter needs to be provisioned with at least 10 MW of cooling capacity to maintain a target inlet temperature.

Our provisioning approach. The traditional provisioning approach guarantees that the datacenter will always be able to support the peak IT power consumption, regardless of outside conditions. This approach makes it impossible to use free cooling in many parts of the world. To see this, Figure 1 shows threshold temperatures around the world, where there is a 2% chance of the threshold or higher temperatures occurring in a year, after the outside air has passed through the evaporative cooling process. We compute these temperatures using ASHRAE's historical temperature distribution data for each location [2], together with a simple well-known thermal model for the evaporative process (described in Section 5). Thus, if one were to target a maximum inlet temperature of $32°C$ in the datacenter at all times, it would be impossible to use free cooling in many parts of the world. Critically, it is important to site datacenters in these geographic areas to provide low response times, since a large fraction of the global population resides in these areas. To reduce costs, we would like to use free cooling there as well.

Several observations suggest that the traditional provisioning approach is overly conservative, and so unnecessarily expensive. First, it is rare for cloud datacenters to operate at (or even near) peak processing capacity [5, 9, 27]. Second, extreme outside temperatures are also rare. As Figure 1 suggests, most locations in the world experience dry-bulb temperatures of 32°C or higher less than 2% of the time. Third, many cloud workloads (*e.g.*, data analytics, scientific computing, interactive workloads other than Web search) may be amenable to a small amount of temporal delay and/or degraded quality of service during the rare times when peak load occurs during overly harsh outside conditions. Finally, even if the workload cannot be appropriately shaped, the IT equipment can operate at higher-than-target temperatures without failing (*e.g.*, [24]), although exposure to high temperatures may increase failure rates over time (*e.g.*, [35]).

Given the above observations, *we propose to underprovision the cooling system based on total cost of ownership (TCO), subject to required performance constraints*, rather than provision for handling peak load under worst-case outside conditions. In our provisioning approach, the datacenter may experience short excursions into temperatures higher than ASHRAE's limits. From the world-wide temperature data, we estimate that allowing temporary excursions of $3°C$ beyond ASHRAE's allowable limits for cloud datacenters is sufficient to allow free cooling with evaporative coolers virtually everywhere on Earth, including most of India and China. With climate change, a few regions eventually may require larger excursions. (If allowed by performance constraints, workload scheduling and power management techniques can significantly reduce these excursions, as we mention below.) We express the cost of these excursions as the additional cost of replacing failed components due to higher temperature. As shall be seen in Section 6, the expected increase in failure rate is small; thus, we expect that current datacenter fault-tolerant software and existing levels of data replication can easily cope with the increased failure rate.

Our approach uses an expected power profile and outside conditions over the course of the datacenter's lifetime. Since it is difficult to predict a detailed power profile so far (*e.g.*, one year or more) in advance, the approach can use a simple abstract expected workload when a detailed profile is not available. Such an abstract profile needs to include a rough estimate of the peaks and valleys of the power consumption over some representative time interval (*e.g.*, one week). This power profile should include the highest expected peak in power consumption that the datacenter will likely experience. The daily peaks of the power profile can be predicted based on the expected peak loads for each day, running on the IT hardware that will be used at the desired level of hardware utilization, whereas the valleys of the power profile can be cast as a percentage (*e.g.*, 33%) of the peaks. We can then connect the daily peaks and valleys using sine waves, and replicate the connected power profile for the planned lifetime of the datacenter. As the cloud provider must forecast its needs (*e.g.*, its expected workload) before building the datacenter, it can be reasonably expected to produce such a simple workload representation. For the outside conditions profile, we use past temperatures and humidities at the datacenter's location, going back as many years as the

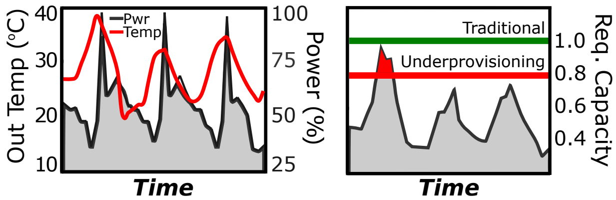

Figure 2. Different provisioning approaches.

planned lifetime of the datacenter. We discuss performance constraints in detail in Section 4.6. Our approach computes the required cooling capacity (in Watts) over time from this resulting power plus outside conditions lifetime profile, and operates on it to provision the cooling infrastructure. The outcome of our approach is the set of cooling technologies (*e.g.*, chillers, free cooling, air conditioning) and respective capacities that should be used to minimize the TCO within the performance constraints.

Figure 2 illustrates three days of a lifetime profile. The left graph in the figure shows the expected daily peak power profile (Y-axis on the right) of the datacenter to be built, as well as the corresponding highest outside temperature (Y-axis on the left) over the last year at the datacenter location, as a function of time (3 days in the X-axis). We do not show humidities for simplicity. The right graph shows the required cooling capacity corresponding to each day depicted in the left graph. This graph also includes two horizontal lines for traditional provisioning (top) and underprovisioning (bottom). As the graph illustrates, traditional provisioning is expensive since peaks of power consumption are rare and they overlap with extreme outside conditions even more rarely. Underprovisioning reduces costs significantly, but may involve rare periods in which there is not enough cooling capacity, *i.e.* the red area in the figure.

In essence, our approach carefully selects where to place the provisioning line, so that the red areas are just large enough to maximize the cost reduction without violating the performance constraints set by the cloud provider. The workload and energy management techniques that our approach prescribes either reduce or eliminate the red areas by lowering power consumption (via consolidation or DVFS) or moving the corresponding load under the line slightly later in time (via deferring the load, if that is acceptable to the provider). The cloud provider can provide information on the techniques' impacts on power and performance, essentially specifying how the techniques impact the required cooling capacity over time. For best results, the datacenter must use these or equivalent techniques when in operation, and can rely on online real-time weather data, *e.g. weather.com*. To compensate for any underprediction of load or power draw, the provider can add slack to the provisioning. If, during operation, the load or power is even higher, the management techniques can control temperature. When such techniques cannot be used, our approach may decide to "rightsize" the cooling, like in the traditional approach.

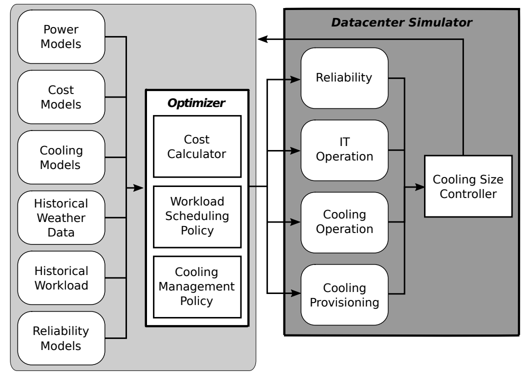

Figure 3. CoolProvision architecture.

Interestingly, our approach may actually leave some amount of red area by design, when the corresponding reduction in cooling CAPEX and OPEX costs exceeds the cost of the additional hardware replacements. Clearly, our approach can lower costs significantly, but it needs to model and quantify these tradeoffs carefully using the lifetime profile described above as its input. CoolProvision embodies these smarts with a default set of models, which the provider can easily replace with more sophisticated ones.

4. CoolProvision

4.1 Overview

CoolProvision proposes a provisioning for the cooling infrastructure that will minimize the cooling-related costs under a set of runtime constraints (inlet air properties and workload performance) by trading off cooling CAPEX and OPEX, IT performance, and IT component reliability. The input to CoolProvision is the power and environmental conditions lifetime profile we introduced above.

Figure 3 illustrates CoolProvision's architecture. It has two main components: a datacenter simulator and a TCO optimizer. The components interact to simulate the entire lifetime of the datacenter, one *time horizon* (6 hours by default) at a time. For finer granularity decisions, we split the horizon into epochs (1 hour each by default). (Large datacenters are typically split into many smaller areas, sometimes called "Colos", so we can simulate just one of these areas and extrapolate from it.) Based on a default starting cooling capacity, configuration, and the input profile, the optimizer decides on the cooling and workload (represented by its power consumption) settings for each epoch in the first horizon. Given these settings, the simulator then simulates the cooling behavior, thermals, power, and workload (represented by the power consumption and CPU utilization). The status of these characteristics after the first horizon is then communicated to the cooling size controller (part of the simulator), which may decide to increase the cooling capacity for the rest of the simulation. This decision impacts the CAPEX of the cooling. The TCO optimizer also receives the status information, selects the settings for each epoch of the next horizon, and invokes the simulator again. This process repeats until all horizons have elapsed and the needed cooling provisioning (technologies and capacities) has been determined. Finally, we rerun the simulation with the produced cooling capacity to obtain accurate energy and reliability results.

The optimizer relies on the power and outside conditions lifetime profile, as well as models to quantify the trade-offs between cooling capital costs, cooling operational costs, and component reliability. Specifically, the optimizer utilizes four classes of models: (1) capital and operating costs, (2) cooling and temperature, (3) power and performance, and (4) IT component reliability. Besides the lifetime profile, the optimizer needs information on the expected impact of the workload and energy management techniques it embodies – workload throttling (via DVFS), workload consolidation onto fewer active servers, and workload deferral (when the latter is allowed) – and how extensively these techniques can be used (*i.e.*, the performance constraints). The techniques' impact is expressed as (1) the change in IT power consumption when the techniques are used, and (2) the change in future IT power. For example, workload deferral can shift power from a load peak to a later time when the underprovisioned cooling infrastructure can handle it. In this case, the performance constraint would specify the maximum number of epochs by which the load peak can be deferred.

With the above inputs and models, the optimizer uses numerical methods, namely sequential quadratic programming (SQP), to minimize the total cooling-related costs within the temperature, relative humidity, and performance constraints defined by the cloud provider. The optimizer's output includes the best cooling speed (amount of airflow and compressor speed), number of servers to keep active, and workload management technique to use for the epochs in the horizon. We define fairly short horizons because solving larger SQP problems can take very long. With shorter horizons, the solver has less future knowledge, which primarily affects load deferring. In practice, providers can define longer horizons and apply more computing power to solve the problem.

4.2 Optimization and Problem Formulation

Table 2 summarizes the input, output, intermediate variables and constants in our framework, whereas Equation 1 states the optimization problem. The framework inputs are: (1) the cooling-related costs (CAPEX, power consumption, and efficiency), (2) the current and future outside conditions, (3) the current conditions inside the datacenter, (4) the IT-related parameters (number of servers, hardware component failure rates), and (5) the optimization-related parameters (*e.g.*, the length of the datacenter lifetime, horizon, and epoch). The outputs are the type and size of the cooling, as well as the settings for each epoch. The settings are: (1) the speed of the cooling (*i.e.*, amount of air flow), (2) the duty cycle of the cooling (on/off time during each epoch), (3) the number of active servers and their corresponding power (DVFS, on/off) state, and (4) an execution plan for the incoming workload in

Symbol Name	Description	Value	Units
Costs			
$TCCost$	Total Cooling-Related Cost	Output	\$
$CoolCapex$	Cooling Capex	Output	\$
$CostPerKw$	Cooling Capex Cost/kW	Input*	\$/W
$CCost_t$	Cooling Opex	Output	\$
$Cost_{power}$	Energy Price/kWh	0.08	\$
$RCost_t$	Reliability Cost	Output	\$
Cooling			
$CoolCapacity$	Cooling Capacity	Int. Var.	kW
$PCool_t$	Cooling Power Draw	Int. Var.	kW
$PComp_t$	Compressor Power	Int. Var.	kW
$PCompMax$	Max Compressor/Chiller Power	{1.6, 5849}	kW
$PFan_t$	Fan Power	Int. Var.	kW
$PFanMax$	Max Fan Power	{0.55, 2025}	kW
$Cool_t$	Cooling Speed	Output	%
E	Evap. Efficiency	90	%
Environmentals			
$Tout_t$	Outside Dry Bulb Temp.	Input	$^\circ C$
$ToutWB_t$	Outside Wet Bulb Temp.	Input	$^\circ C$
Tin_t	Inlet Temp.	Int. Var.	$^\circ C$
RH_t	Inlet RH	Int. Var.	%
$TinVar_t$	Inlet Temp. Variation/hour	Int. Var.	$^\circ C$
$RHinVar_t$	Inlet RH Variation/hour	Int. Var.	%
Tin_{Low}	Inlet Temp. Low Bound	5	$^\circ C$
Tin_{Set}	Inlet Temp. Setpoint	30	$^\circ C$
Tin_{Tol}	Inlet Temp. Tolerance	5	$^\circ C$
RH_{High}	RH High Bound	80	%
RH_{Low}	RH Low Bound	20	%
$RHVar_{High}$	Max RH Variation/hour	20	%
$TVar_{High}$	Max Temp. Variation/hour	20	$^\circ C$
$TCase_t$	Disk Temp.	Int. Var.	$^\circ C$
Servers			
Max_{srv}	Total Servers	{64, 40000}	N/A
f_t	CPU Frequency	Output	%
L_t	IT Power as % of Max Power	Output	%
$PDyn_t$	Dynamic Power	Int. Var.	W
$PLeak_t$	Leakage Power	Int. Var.	W
$P_{leak\_nominal}$	Leakage at 40°C	{7, 80}	W
$TCPU_t$	CPU Temperature	Int. Var.	$^\circ C$
$PMax$	Server Max Power	{35, 250}	W
$PBack$	Server Background Power	{13, 45}	W
$NodePower_t^k$	Server Power	Int. Var.	W
Disks			
$Disk_{cnt}$	Disk Count	{64, 160000}	N/A
$Disk_{cost}$	Disk Cost	100	\$
AFR	Disk AFR at 40°C	1.5	%
$Pfail_t$	Disk Fail Rate/epoch	Int. Var.	N/A
$P_{fail\_nominal}$	Disk Fail Rate/epoch at 40°C	Int. Var.	%
CoolProvision			
l	Epochs in Lifetime	15×8760	N/A
h	Horizon	6	N/A
t	Current Epoch	$0-l$	N/A
$\|t\|$	Epoch Length	60	Mins

Table 2. Summary of the framework. Each symbol is categorized as an input, output, or intermediate variable for the optimization. Default values are shown whenever possible and should be regarded as inputs. Wherever there are two input values, the one on the left is used in our study of small datacenters and the one on the right is for large datacenters. *For *CostPerKw* we use values from Parasol for small datacenters (3.2 \$/W for the direct expansion AC and 0.96 \$/W for the direct evaporative cooler), and values from Table 1 for large datacenters.

case it is deferrable. Note that, for simplicity, we focus solely on the reliability of disk drives, since these components are the most likely to be affected by a harsh environment [34].

$$
\begin{aligned}
&\text{min: } TCCost = \forall\, t \in \{0, h, 2h, \ldots, \frac{l}{h}\} \\
&\qquad \sum_t (CoolCapex \cdot \frac{h}{l} + \sum_t^{t+h} (CCost_t + RCost_t)) \\
&\text{s. t.: } Tin_{Low} \leq Tin_t \leq Tin_{Set} + Tin_{Tol} \\
&\quad RH_{Low} \leq RH_t \leq RH_{High} \\
&\quad TinVar_t \leq TVar_{High} \\
&\quad RHinVar_t \leq RHVar_{High}
\end{aligned}
\tag{1}
$$

CoolProvision's optimization goal is to minimize the cost of purchasing (*CoolCapex*) and operating (*CCost*) the cooling infrastructure, and replacing failed hardware (*RCost*) over the horizon. *CoolCapex* includes the capital cost of the cooling equipment, installation, and piping, whereas *CCost* represents the energy cost of operating this equipment.

The constraints relate to environmental conditions within the datacenter. Specifically, the first constraint in Equation 1 refers to keeping inlet air temperature (Tin) between a minimum ($TinLow$) and a setpoint (T_{Set}), while allowing a tolerance (T_{Tol}) if keeping the temperature at the setpoint is not feasible. The redline temperature we absolutely do not want to exceed is $Tin_{Set} + Tin_{Tol}$. The second constraint limits the inlet air relative humidity within a range (RH_{Low}–RH_{High}). The third and fourth constraints limit the amount of inlet temperature and relative humidity variation.

CoolProvision solves the optimization problem as many times as there are horizons in the projected lifetime of the datacenter. If any of these solutions is infeasible, CoolProvision applies one of the workload and energy management techniques to the abstract power profile, and re-starts the procedure for that horizon. The techniques modify the power profile based on the performance constraints specified by the cloud provider (discussed below). As performance constraints, the provider can specify the acceptable amount of DVFS during a year (*i.e.*, the percentage of time DVFS can be used), the amount of consolidation (*i.e.*, the ratio between the number of active servers with and without consolidation), and the amount of workload deferral (*i.e.*, the number of epochs by which the load can be deferred) that is allowed.

4.3 Cost Modeling

Cooling capital cost. We model this cost as: $CoolCapex = CoolCapacity \cdot CostPerKW$, where $CoolCapacity$ is the capacity of the cooling system measured in KW, and $CostPerKW$ is the capital cost of the cooling infrastructure per KW.

We assume that the cooling capacity/size has a linear impact on the peak cooling power. To characterize this relationship, we have collected data (cooling capacity-to-power ratio) from several vendors for four air and water-cooled chillers and one air-side economizer with backup chiller (75 models in total). Our collected data suggest that the ratio remains approximately constant as each technology scales towards larger sizes, and justify our assumption.

We also assume that the capacity/size of the cooling equipment (fans, compressors, pipes, sensors, and ducts) scales linearly with their price, including deployment, installation, and maintenance. We verified the accuracy of this assumption with data from several vendors.

Cooling operational cost. The cooling OPEX cost during epoch t depends on the average power consumption of the cooling equipment $\overline{PCool_t}$ and the energy price $Cost_{power}$: $CCost_t = \overline{PCool_t} \cdot Cost_{power}$.

Reliability cost. Disks account for a large fraction (*e.g.*, 75%) of failures in datacenters, while CPU and memory each accounts for less than 5% [32, 34, 37]. Thus, we model the reliability cost as that for replacing disks that will likely fail due to higher temperature; our model can be easily extended to account for other component failures as well. Specifically:

$$RCost_t = |t| \cdot (Pfail_t - P_{fail\_nominal}) \cdot Disk_{cost} \cdot Disk_{cnt} \tag{2}$$

where $|t|$ is the duration of epoch t, $Pfail_t$ is the disk failure rate during the epoch, $P_{fail\_nominal}$ is the fraction of the Annualized Failure Rate (AFR) during the epoch, $Disk_{cost}$ is the disk price, and $Disk_{cnt}$ is the number of disks.

To compute $Pfail_t$, we use the Arrhenius model to calculate an *acceleration factor* (AF_t) for the disk lifetime as a function of temperature. This model has been used for the same purpose in prior works, *e.g.* [11, 34]. AF_t determines how fast temperature accelerates the disks' failure:

$$AF_t = e^{\frac{E_a}{k} \cdot (\frac{1}{\overline{TCase_t}} - \frac{1}{\overline{T}})} \tag{3}$$

where E_a is the activation energy (in eV) for the device, k is Boltzmann's constant ($1.1 \cdot 10^{-5}$ eV/K), $\overline{T}$ is the baseline operating temperature (in K) of the device, and $\overline{TCase_t}$ is the average elevated temperature (in K) of the device. Note that this model makes the reliability cost become a non-linear function of temperature. We then compute $Pfail_t$ as:

$$Pfail_t = AF_t \cdot P_{fail\_nominal} \tag{4}$$

For example, for a temperature of $60°C$, Equations 3 and 4 predict a two-fold increase in the failure probability.

4.4 Thermal and Cooling Modeling

Inlet air temperature and relative humidity. CoolProvision requires models that can predict the inlet air temperature and relative humidity at each server. These models are typically functions of the following inputs: (1) the current inlet temperature and humidity, (2) the outside temperature and humidity, (3) the current IT load (power consumption), (4) the cooling unit that is operating, and (5) the speed at which it is operating (*e.g.*, fan, compressor).

These models should be provided by the cloud provider for each cooling technology. In Section 5, we describe the specific models we use in our Parasol case study.

Disk case temperature. This temperature can be approximated as a linear function of the inlet temperature, CPU utilization, and disk utilization. For simplicity, we conservatively set it to the worst-case hot aisle temperature.

4.5 Power Modeling

Server power model. We model the power consumption *NodePower* of the k^{th} server as a function of the CPU utilization (which we estimate from the workload simulation), temperature, and frequency. We split the server power into: (1) dynamic CPU power $PDyn_t$, (2) CPU leakage power $PLeak_t$ and (3) background power from the other components *PBack*. As in [36], we model dynamic power as a linear function of frequency and utilization (assuming that the power draw of the other components varies little). We also assume that all cores share the same voltage and frequency. A sleeping server (*e.g.*, ACPI S3) consumes near-zero Watts.

$$NodePower_t = \overline{PLeak_t} + \overline{PDyn_t} + PBack \tag{5}$$

$$PDyn_t = a \cdot f_t \cdot \overline{CPUutil_t} \cdot PMax \tag{6}$$

For the leakage power, we use a model from Biswas *et al.* [8] that relates temperature and leakage. Assuming an operating temperature of $\overline{TCPU_t}$ that matches the inlet air temperature, we estimate the leakage as a function of a known base leakage $P_{leak\_nominal}$ at $TCPU = 40°C$.

$$PLeak_t = (a \cdot \overline{TCPU_t}^2 + b \cdot \overline{TCPU_t} + c) \cdot P_{leak\_nominal} \tag{7}$$

Cooling power model. CoolProvision requires models relating the speed of operation of each cooling unit to its power consumption. The cloud provider must provide these models for the cooling technologies to be considered. In Section 5, we detail the power models we use in our Parasol case study.

4.6 Workload and Energy Management

Using the models described in Sections 4.4 and 4.5 and the input (IT power plus outside conditions) lifetime profile, CoolProvision can determine the required cooling capacity over time. However, provisioning for the maximum requirement can be expensive, so it also embodies workload and energy management techniques – workload throttling (via DVFS), workload consolidation onto fewer active servers, and workload deferral (when the latter is allowed) – to lower the higher cooling capacity requirements. For best results, these techniques should be used by the datacenter management software when the datacenter starts to operate.

Instead of applying the techniques to detailed workload traces that are unlikely to be accurate so far in advance of operation, CoolProvision evaluates the techniques analytically in terms of their impact on the IT power profile. The cloud provider specifies which techniques are acceptable, based on the characteristics of the expected workload (*e.g.*, batch vs interactive). The techniques change the power draw when they are used, and possibly the future power draw when using them causes load to be postponed or lengthened. The provider must also specify the extent of these effects on the power profile, and limits on the techniques' use (*e.g.*, do not DVFS the CPU for more than 5% of the epochs, do not exceed 80% CPU utilization when consolidating). We assume that the provider knows the limits its workloads would be under. We refer to these limits as the performance constraints.

Our algorithm for applying the techniques works as follows. It starts when CoolProvision finds Equation 1 to be infeasible (usually because the inlet temperature becomes higher than $Tin_{Set} + Tin_{Tol}$) for a horizon. At that point, the algorithm performs a gradient descent by repeatedly changing the power profile according to the technique used (and within its performance constraints), and evaluating the equation for the horizon. Currently, CoolProvision uses only one technique per optimization run. Eventually, CoolProvision finds the cheapest setting of the power profile that will satisfy all constraints. Next, we discuss each technique in turn.

Workload Deferral. Workload deferral often applies to batch computations that do not have tight deadlines. This technique may allow CoolProvision to shift some of the expected IT energy consumption (heat production) to times when there is less load in the system or when cooling is more efficient. For example, free cooling is most efficient when both humidity and temperature are relatively low. CoolProvision guarantees that the workload cannot be deferred more than a provider-chosen number of epochs.

Workload Consolidation. Both batch and interactive workloads can often be consolidated onto fewer servers, so that the freed-up servers can be transitioned to a low-power state (*e.g.*, ACPI S3). This technique allows CoolProvision to lower the expected IT power draw when the outside environmental conditions are expected to be harsh (*e.g.*, noon time during the summer at a warm location) or when power draw is expected to be very high. To apply this technique, CoolProvision requires two additional inputs from the cloud provider: the typical CPU utilization at the highest expected load peak, and how much higher is CPU utilization of the active servers is allowed to get after consolidation. These two values determine the maximum amount of consolidation that the cloud provider wants to allow. For example, an interactive workload may target a peak CPU utilization of 50%, (response time often suffers at higher utilization due to queuing effects), but may accept that a subset of the servers run at 60% utilization during short periods of extreme conditions. To calculate the power draw of the consolidated cluster, the algorithm assumes that the load on the servers to be deactivated is evenly spread (*e.g.*, via live VM migration or request distribution) across the servers that will remain active.

Workload Throttling via DVFS. Workload throttling can usually be applied to both batch and interactive workloads.

Similar to the two previous techniques, throttling can reduce the required cooling capacity by "shaving" the thermal peaks. CoolProvision assumes that all servers are throttled together and by the same amount. When applying this technique, CoolProvision searches for the frequency that will minimize the cooling-related cost. To do this, CoolProvision requires three inputs from the cloud provider: (1) a function quantifying power consumption for each allowed DVFS setting during load peaks, (2) a function quantifying performance loss (or how much the power profile should be shifted to the right) for each DVFS setting during load peaks, and (3) a constraint on the percentage of epochs during a year where CoolProvision can apply DVFS.

5. Applying CoolProvision to Parasol

We do not have unfettered access to a cloud datacenter and its cooling, so we demonstrate CoolProvision with models developed for Parasol in [19]. Parasol is a small container-based datacenter that combines a free cooling unit (Dantherm Flexibox 4500) and a direct-expansion AC (Dantherm iA/C 19000). It also embodies a commercial cooling controller that periodically selects the cooling regime (*e.g.*, free cooling at 50% fan speed, or free cooling off with AC on). To broaden the geographical regions where Parasol could be efficiently used, we model an additional washer system in Parasol's free cooling unit to lower air temperature via evaporation. With the additional model, Parasol becomes a hybrid direct evaporative cooler plus an AC.

Using the Parasol models and the workload and energy management techniques we consider, CoolProvision can determine what an underprovisioned cooling system for Parasol could have been and how much money it would save. Similarly, we can use these models and the management techniques to provision a different datacenter with one or both of the cooling technologies we consider.

Goiri *et al.* built the Parasol thermal and cooling models using regression over several months of temperature, humidity, and power consumption data [19]. Next, we describe the models that we use for the free cooling unit and the AC. We also leverage Parasol's free cooling model to derive a model for an evaporative cooler (which Parasol does not have).

Compressor-based AC thermal model. We model a variable-speed compressor as in [19]. The model predicts the next inlet air temperature Tin_{t+1} based on several factors:

$$Tin_{t+1} = \alpha \cdot Cool_t + \beta \cdot Tout_t + \gamma \cdot Tin_t + \delta \cdot Tin_{t-1} + \varepsilon \cdot L_t + \zeta \cdot Cool_t + \eta \quad (8)$$

where $Cool_t$ is the cooling speed, L_t is the percentage of the maximum datacenter power, and $\alpha \ldots \eta$ are constants estimated using regression from Parasol's AC unit. We model the relative humidity as a constant when using the AC.

Direct evaporative cooler thermal model. We use the theoretical model for an evaporative cooler [1] to alter the properties of the outside air, effectively simulating the decrease in dry bulb and the increase in wet bulb temperature due to water evaporation. The relationship is as follows:

$$ToutNewDB_{t+1} = Tout_{t+1} - ((Tout_{t+1} - ToutWB_{t+1}) \cdot E) \quad (9)$$

where $ToutNewDB_{t+1}$ denotes the new dry bulb temperature, $Tout_{t+1}$ and $ToutWB_{t+1}$ denote the original dry and wet bulb temperatures, and E denotes the efficiency of the medium. To compute the inlet temperature, we set $Tout_{t+1} = ToutNewDB_{t+1}$. To compute the relative humidity, we operate with the dry and wet bulb temperatures using a model from [4]. Then, we use the new air properties as an input to the existing Parasol free cooling model:

$$Tin_{t+1} = \alpha \cdot Cool_{t-1} + \beta \cdot Cool_t \cdot Tout_t + \gamma \cdot Cool_{t-1} \cdot Tin_{t-1} + \delta \cdot Tout_t + \varepsilon \cdot Tin_{t-1} + \zeta \cdot Tin_{t-1} + \eta \cdot L_t \quad (10)$$

When it is too cold outside, the optimizer lowers the fan speed (and may even close out the datacenter) to increase the temperature inside. When the outside relative humidity is higher than RH_{High}, the optimizer re-circulates the warm air to raise temperature (thus lowering relative humidity).

AC power model. We model the AC power as the sum of the compressor (*PComp*) and fan (*PFan*) powers.

$$PComp_t = \alpha \cdot Cool_t \cdot PCompMax + \beta \quad (11)$$

$$PFan_t = (\gamma \cdot (Cool_t^2) - \delta \cdot Cool_t) \cdot PFanMax + \varepsilon \quad (12)$$

where *PCompMax* and *PFanMax* are the maximum power draws of the AC compressor and fan, respectively.

Direct evaporative cooler power model. We model this power draw as $PFan_t$ above but with other constants instantiated from measurements of Parasol's free cooling unit [19]. We assume that the power draw of the washer is negligible.

6. Evaluation

6.1 Methodology

We evaluate CoolProvision in two main scenarios. First, we use it to investigate how much we could have saved by underprovisioning Parasol, as an example of a small-scale datacenter. Then, we re-target CoolProvision to explore larger datacenters. In both scenarios, we seek to determine the cooling capacity that will minimize the cooling-related costs, while respecting all environmental and performance constraints, according to Equation 1.

We drive CoolProvision using a power plus outside conditions lifetime profile constructed as described in Section 3. To study realistic profiles, we build them based on the characteristics of realistic workload power traces. Table 3 lists the characteristics of our traces, the techniques we allow CoolProvision to apply to them, and their corresponding performance constraints. (We scale the traces linearly to keep the same average utilization and peak-to-valley load ratio across small and large datacenters.) The Fixed-High traces

Power Trace	Management Technique	Performance Constraint	Average Utilization	Peak/Valley Load Ratio
Fixed-High-1	None, rightsized	N/A	100%	1.00
Fixed-High-2	None, underprov.	N/A	100%	1.00
Batch	Throttling	5% per year	72%	8.00
Batch-Defer	Deferral	6 hours	72%	8.00
Interactive-1	Throttling	5% per year	57%	2.90
Interactive-2	Consolidation	62% ratio	52%	4.92

Table 3. Power traces and management techniques.

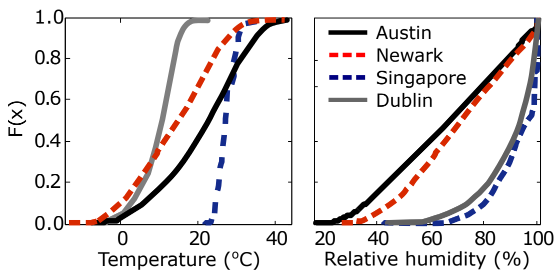

Figure 4. Environmental conditions at our locations.

represent the traditional approach of provisioning based on a single maximum power value (100% cluster utilization). In the Fixed-High-1 results, we do not allow CoolProvision to underprovision the cooling infrastructure. In the Fixed-High-2 (also peak utilization) results, we allow CoolProvision to underprovision the cooling infrastructure, but not to use any workload/energy management techniques. Fixed-High-2 directly trades off savings in cooling cost against the cost of hardware replacements due to harsh environmentals. The Batch power traces are based on a Facebook trace [10], whereas Interactive-1 and Interactive-2 are based on Hotmail [38] and Ask.com [3] traces, respectively. The Batch-Deferrable results allow the workload to be deferred by up to 6 hours. In the throttling cases, we only allow CoolProvision to DVFS the CPUs at most 5% of the epochs in each year. The CPUs can be throttled to half of their peak frequency in increments of 100MHz, with power consumption dropping by 55% at the slowest speed. Finally, we allow consolidation in Interactive-2 down to 62% of active servers. In the absence of real performance constraints, we use these plausible values to demonstrate the operation of CoolProvision.

For our results, we run CoolProvision while specifying no performance loss for doing workload throttling or consolidation, and the performance constraints listed in Table 3. The reason is that, given these constraints and the target utilization, the techniques produce losses that are insignificant compared to our default epoch length (60 minutes).

In terms of outside environmental conditions, we study four locations with different weather patterns: (1) hot and dry with low temperature variation (Austin), (2) hot and dry with high temperature variation (Newark), (3) hot and humid with low temperature variation (Singapore) and (4) cold and humid with low temperature variation (Dublin). Figure 4 shows the main characteristics of the locations using Cumulative Distribution Functions (CDFs) of outside temperature (temperature) and outside relative humidity (right).

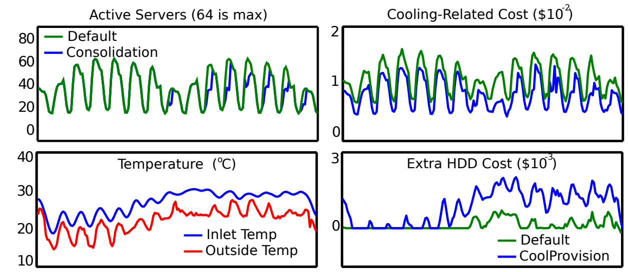

Figure 5. Behaviors over 15 days in June for a small-scale datacenter with free cooling *(X-axes are time)*.

We consider three cooling technologies: direct-expansion AC and direct evaporative (free) cooling for small datacenters, and chiller and free cooling for large datacenters. Although CoolProvision is capable of provisioning the cooling infrastructure with multiple technologies, we only consider infrastructures with one technology in this paper.

For each technology and location pair, we use a different baseline cooling capacity. To obtain it, we configure CoolProvision (using Fixed-High-1 as the power trace) to identify the minimum capacity that keeps (1) $T_{Set} = 30°C$ without workload/energy management for the AC and chiller systems, and (2) the closest temperature to $T_{Set} = 30°C$ for the evaporative cooler without workload/energy management. *These baselines mimic the traditional provisioning.*

6.2 Underprovisioning small datacenters

Our first evaluation study applies CoolProvision to Parasol. For this study, we instantiate the constants in our models using real measurements from Parasol. The Parasol custom container currently hosts 64 low-power (25-30W) servers organized in a single cold aisle.

Figure 5 shows an example of the behaviors that CoolProvision projects for two weeks of Interactive-2 with direct evaporative cooling in Newark. The figure shows the number of active servers (top left), the total cooling-related cost per hour (top right), the inlet and outside air temperatures (bottom left), and the disk replacement cost per hour (bottom right). Note that the times when CoolProvision prescribes consolidation are exactly the times with both high power consumption and outside temperatures.

Figure 6 depicts our cost and capacity results for all lifetime profiles and locations. On the left of the figure, we present the cooling-related costs and the capacity requirements for Parasol's AC. On the right, we present the equivalent results for an evaporative cooler. The bars in each group can only be directly compared to Fixed-High-1, the traditional provisioning, since they differ in performance constraints and/or expected workloads.

Figure 7 shows the inlet air temperature CDFs for the three warmest locations under AC and free cooling. We do not show CDFs for Dublin, but the trends are similar.

Next, we summarize our findings from these figures:

(1) CoolProvision produces provisionings that are substan-

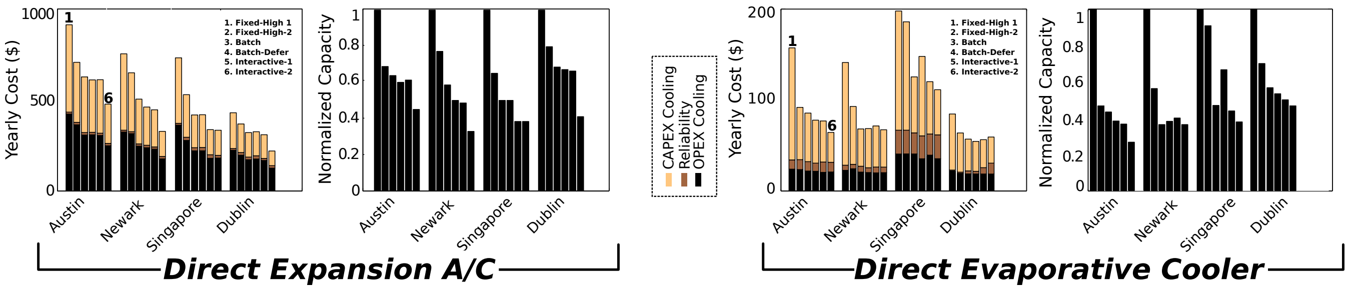

Figure 6. Yearly cooling-related cost and normalized cooling capacity when provisioning a small datacenter with a direct-expansion AC (left) and an evaporative cooler (right) at four locations.

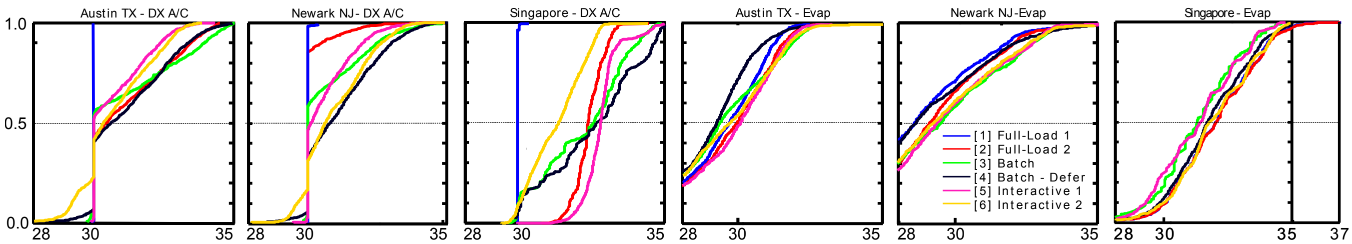

Figure 7. Inlet temperature CDFs for a direct expansion AC and a direct evaporative cooler at three locations.

tially cheaper than traditional provisioning, due to significant reductions in required cooling capacity. Underprovisioning reduces both CAPEX and OPEX costs, regardless of cooling technology. One can see these effects by comparing the bars for Fixed-High-1 to the other bars in Figure 6.

(2) As expected, the location heavily affects the cooling-related costs. AC-based cooling is most expensive in Austin whereas free cooling is most expensive in Singapore. Dublin shows the lowest costs regardless of technology.

(3) Underprovisioning without workload/energy management (Fixed-High-2) reduces costs and required capacities compared to traditional provisioning (Fixed-High-1). For example, it reduces required capacities by 21-32% for the AC system, and even more under free cooling. Despite its low reliability cost, CoolProvision cannot underprovision Fixed-High-2 further, due to the limit at $35°C$ (Figure 7).

(4) Allowing CoolProvision to use workload/energy management produces even greater savings. Although not comparable to each other, the management policies achieve cost reductions of 21-55% for the AC system and 20-43% for the free cooling system, compared to traditional provisioning.

(5) The policies may even enable a cheaper technology. Fixed-High-1 and Fixed-High-2 exceed $35°C$ when using free cooling in Singapore. But when CoolProvision can use the policies, temperatures always stay below the limit. This shows that an AC need not be used in Singapore, if free cooling is combined with workload/energy management.

(6) Workload throttling is most effective for locations with relatively rare outside temperature peaks (*e.g.*, Newark), because it can easily shave these peaks in required cooling capacity. Consolidation enables the same kind of shaving but for longer periods, so it is consistently effective.

(7) Under AC cooling, workload deferral is most effective for locations where outside temperature variation is high, since it can defer the load until the outside temperature decreases and cooling efficiency improves. Under free cooling, workload deferral also allows shifting the heat production to drier times of the day, which is useful in Singapore. Allowing deferrals of up to 6 hours is enough to shave daily peaks of temperature and relative humidity.

(8) The reliability cost is a small percentage of the overall cooling-related cost, regardless of technology. However, the Interactive-1 and Interactive-2 results for Dublin under free cooling show that, in some scenarios, the increase in reliability cost may prevent further cost reductions.

(9) CoolProvision raises the cost of replacing disks by increasing the expected AFR from 1.5% to 1.63-2.35%. The highest reliability penalties occur in Singapore, regardless of technology, where the resulting inlet temperatures are almost always higher than $30°C$.

(10) Traditional provisioning keeps the inlet temperature almost always at $30°C$ under the AC system, whereas the underprovisioned scenarios aggressively exploit the tolerance up to $35°C$. In some cases, temperatures are always higher than the setpoint in Singapore. When using free cooling, the percentage of inlet temperatures that are higher than the setpoint increases significantly when we underprovision. Recall that, for the free cooling system, our baseline seeks to keep the temperature as close as possible to the setpoint.

6.3 Underprovisioning large datacenters

Our second evaluation considers a hypothetical large 10MW datacenter cooled via water chillers or a direct evaporative cooler. To mimic this datacenter, we model our small datacenter as a single cold/hot aisle, and then replicate this aisle to form the large datacenter. Admittedly, this extrapolation is only a rough approximation, but should provide a sense for the possible underprovisioning savings in large datacenters. We also use different input parameters in some cases (see Table 2). For example, we replace Parasol's cooling CAPEX costs with those from Table 1; we also adjust the maximum power consumption of the cooling technologies (chillers and direct evaporative cooling) according to PUEs scaled from Table 1; and we alter the servers' characteristics to become closer to those of the more power-hungry servers of large datacenters (*e.g.*, 4-disk Microsoft Cloud server [28]).

Figure 8 shows our main results for the same locations, power traces (scaled to match the larger datacenter size), and workload/energy management techniques as before. The cooling-related cost and required capacity trends are very similar to those of Figure 6. Though we omit the temperature CDFs to save space, we find similar trends there as well. An interesting observation is that the free cooling system cannot maintain even $35°C$ in Singapore. At this location, the higher IT power consumption we assume for this datacenter demands a more expensive cooling technology.

7. Related Work

To our knowledge, this is the *first* paper to explore the underprovisioning of datacenter cooling systems. A few prior works (*e.g.*, [7]) studied cooling rightsizing in older datacenters (lower inlet temperatures, no hot aisle containment, no free cooling) using simulations of heat recirculation. These works did not model multiple cooling technologies, costs, workloads, performance, or reliability. Moreover, heat recirculation is much less relevant in modern datacenters with hot aisle containment.

Our work relates peripherally to prior research on energy-aware dynamic thermal management, dynamic power and thermal emergency management, and reliability modeling.

Dynamic power and thermal emergency management. Heath *et al.* [22] and Ramos *et al.* [33] proposed workload and power management techniques for handling datacenter thermal emergencies, i.e. periods when not enough cooling capacity is available due to cooling equipment failures. Similarly, Fan *et al.* [17] and Govindan *et al.* [21] proposed to underprovision the datacenter power delivery infrastructure, and use power management techniques during power emergencies, *i.e.* periods when power demand exceeds the capability of the provisioned delivery infrastructure.

Though CoolProvision is a cooling infrastructure underprovisioning framework, it does prescribe workload and energy management techniques that should be used during datacenter operation when the provisioned cooling is insufficient. In fact, it is those techniques and CoolProvision's modeling of reliability costs that enable the underprovisioning in the first place. The management techniques we explore in this paper were inspired by the prior works above.

Energy-aware dynamic thermal management. The literature on this topic is extensive, *e.g.* [6, 26, 29, 30, 41]. Most recently, Goiri *et al.* [19] and Kim *et al.* [25] used workload placement and scheduling techniques to efficiently manage temperature and humidity in hybrid (free cooling plus backup cooling) datacenters. These works leverage some of the same workload management techniques (consolidation, deferral) embodied in CoolProvision.

Reliability modeling in (chiller-based) datacenters. Sankar *et al.* [34] indicated an exponential relationship between operating temperatures and hard disk lifetime in datacenters, using the Arrhenius formula. In contrast, El-Sayed *et al.* [12] suggested that the exponential model is pessimistic and a linear relationship may be more representative. They also suggested that variability in operating temperature may be an equally important or even more important factor for disk reliability than the actual operating temperatures.

CoolProvision establishes both temperature and temperature variation limits as constraints in its cooling equipment provisioning. It also models disk reliability using the Arrhenius formula, but a different model can easily be used.

8. Conclusions

In this paper, we proposed an approach for systematically underprovisioning the cooling infrastructure of cloud datacenters. Based on this approach, we introduced CoolProvision, an optimization and simulation framework for selecting the best provisioning with respect to the datacenter's cooling-related cost. To demonstrate CoolProvision's use, we applied it to a real small free-cooled datacenter, and found that it can produce cooling-related cost savings of up to 55%. We also extrapolated our experience and results to larger cloud datacenters and saw similar trends.

Acknowledgements

We thank Dave Gauthier, Kris Johnson, Gregg McKnight, and Steve Solomon for comments that helped improve this paper. We also thank Microsoft for its financial support via a 2014 Software Engineering Innovation Foundation award.

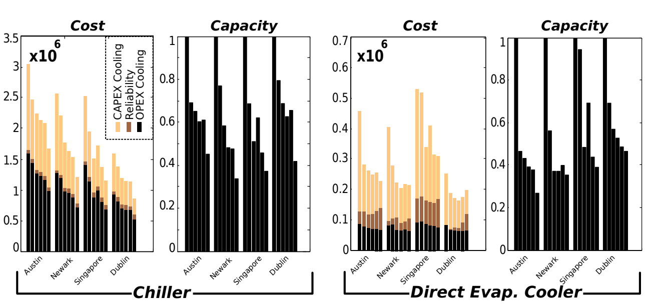

Figure 8. Costs and capacities for a 10MW datacenter.

References

[1] ASHRAE. HVAC Systems and Equipment, 2012.

[2] ASHRAE. ASHRAE Climate Data Center, 2015.

[3] Ask.com. Fraction of Requests from April 2008, 2008.

[4] A. Bahadori, G. Zahedi, S. Zendehboudi, and K. Hooman. Simple Predictive Tool to Estimate Relative Humidity using Wet Bulb Depression and Dry Bulb Temperature. *Applied Thermal Engineering*, 50(1):511–515, 2013.

[5] L. A. Barroso, J. Clidaras, and U. Holzle. *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines*. Second edition, 2013.

[6] C. Bash and G. Forman. Cool Job Allocation: Measuring the Power Savings of Placing Jobs at Cooling-Efficient Locations in the Data Center. In *USENIX ATC*, 2007.

[7] A. Beitelmal and C. D. Patel. Thermo-Fluids Provisioning of a High Performance High Density Data Center. *Distributed and Parallel Databases*, 21(2-3), 2007.

[8] S. Biswas, M. Tiwari, T. Sherwood, L. Theogarajan, and F. T. Chong. Fighting Fire With Fire: Modeling the Datacenter-Scale Effects of Targeted Superlattice Thermal Management. In *ISCA*, 2011.

[9] M. Carvalho, W. Cirne, F. Brasileiro, and J. Wilkes. Long-Term SLOs for Reclaimed Cloud Computing Resources. In *SoCC*, 2014.

[10] Y. Chen, S. Alspaugh, A. Ganapathi, R. Griffith, and R. Katz. Statistical Workload Injector for MapReduce, 2012.

[11] G. Cole. Estimating Drive Reliability in Desktop Computers and Consumer Electronics Systems. *Seagate Technology Paper TP*, (338), 2000.

[12] N. El-Sayed, I. Stefanovici, G. Amvrosiadis, A. Hwang, and B. Schroeder. Temperature Management in Data Centers: Why Some (Might) Like it Hot. In *SIGMETRICS*, 2012.

[13] Emerson Network Power. Energy Efficient Cooling Solutions for Data Centers, 2007.

[14] H. Endo, H. Kodama, H. Fukuda, T. Sugimoto, T. Horie, and M. Kondo. Effect of Climatic Conditions on Energy Consumption in Direct Fresh-air Container Data Centers. In *IGCC*, 2013.

[15] T. Evans. The Different Technologies for Cooling Data Centers. *APC white paper*, 59, 2012.

[16] Facebook. Prineville Data Center – PUE/WUE, 2014.

[17] X. Fan, W. Weber, and L. Barroso. Power Provisioning for a Warehouse-sized Computer. In *ISCA*, 2007.

[18] Í. Goiri, W. Katsak, K. Le, T. D. Nguyen, and R. Bianchini. Parasol and GreenSwitch: Managing Datacenters Powered by Renewable Energy. In *ASPLOS*, 2013.

[19] Í. Goiri, T. D. Nguyen, and R. Bianchini. CoolAir: Temperature- and Variation-Aware Management for Free-Cooled Datacenters. In *ASPLOS*, 2015.

[20] Google. Efficiency: How We Do It, 2014.

[21] S. Govindan, D. Wang, A. Sivasubramaniam, and B. Urgaonkar. Aggressive Datacenter Power Provisioning with Batteries. *ACM Transactions on Computer Systems*, 31(1), 2013.

[22] T. Heath, A. P. Centeno, P. George, L. Ramos, Y. Jaluria, and R. Bianchini. Mercury and Freon: Temperature Emulation and Management for Server Systems. In *ASPLOS*, 2006.

[23] Hewlett Packard. Applying 2011 ASHRAE data center guidelines to HP ProLiant-based facilities. Technical report, 2012.

[24] Intel Corp., Microsoft Corp. Server Power and Performance Evaluation in High-Temperature Environments, 2012.

[25] J. Kim, M. Ruggiero, and D. Atienza. Free Cooling-Aware Dynamic Power Management for Green Datacenters. In *HPCS*, 2012.

[26] L. Li, C. Liang, J. Liu, S. Nath, A. Terzis, and C. Faloutsos. Thermocast: A Cyber-Physical Forecasting Model for Data Centers. In *KDD*, 2011.

[27] H. Liu. A Measurement Study of Server Utilization in Public Clouds. In *DASC*, 2011.

[28] Microsoft. Open Compute Project Server Specs.

[29] J. Moore, J. Chase, and P. Ranganathan. Weatherman: Automated, Online and Predictive Thermal Mapping and Management for Data Centers. In *ICAC*, 2006.

[30] T. Mukherjee, A. Banerjee, G. Varsamopoulos, S. Gupta, and S. Rungta. Spatio-Temporal Thermal-aware Job Scheduling to Minimize Energy Consumption in Virtualized Heterogeneous Data Centers. *Computer Networks*, 53(17), 2009.

[31] J. Niemann, J. Bean, and V. Avelar. Economizer Modes of Data Center Cooling Systems. *Schneider Electric Data Center Science Center Whitepaper*, 2011.

[32] E. Pinheiro, W.-D. Weber, and L. A. Barroso. Failure Trends in a Large Disk Drive Population. In *FAST*, 2007.

[33] L. Ramos and R. Bianchini. C-Oracle: Predictive Thermal Management for Data Centers. In *HPCA*, 2008.

[34] S. Sankar, M. Shaw, and K. Vaid. Impact of Temperature on Hard Disk Drive Reliability in Large Datacenters. In *DSN*, 2011.

[35] S. Sankar, M. Shaw, K. Vaid, and S. Gurumurthi. Datacenter Scale Evaluation of the Impact of Temperature on Hard Disk Drive Failures. *ACM Transactions on Storage*, 9(2), 2013.

[36] K. Shen, A. Shriraman, S. Dwarkadas, X. Zhang, and Z. Chen. Power Containers: An OS Facility for Fine-Grained Power and Energy Management on Multicore Servers. In *ASPLOS*, 2013.

[37] The Green Grid. Data Center Efficiency and IT Equipment Reliability at Wider Operating Temperature and Humidity Ranges, 2012.

[38] E. Thereska, A. Donnelly, and D. Narayanan. Sierra: Practical Power-Proportionality for Data Center Storage. In *EuroSys*, 2011.

[39] Uptime Institute. Data Center Industry Survey Results, 2014.

[40] U.S. Environmental Protection Agency. EPA Report on Server and Data Center Energy Efficiency. August 2007.

[41] R. Zhou, Z. Wang, C. B. E, A. McReynolds, C. Hoover, R. Shih, N. Kumari, and R. Sharma. A Holistic and Optimal Approach for Data Center Cooling Management. In *ACC*, 2011.