Thermal Control Strategies for Reliable and Energy-Efficient Data Centers

Rehan Khalid[1]
Department of Mechanical Engineering,
Villanova University,
800 Lancaster Avenue,
Villanova, PA 19085
e-mail: rkhalid@villanova.edu

Aaron P. Wemhoff
Department of Mechanical Engineering,
Villanova University,
800 Lancaster Avenue,
Villanova, PA 19085

Two self-developed control schemes, ON/OFF and supervisory control and data acquisition (SCADA), were applied on a hybrid evaporative and direct expansion (DX)-based model data center cooling system to assess the impact of controls on reliability and energy efficiency. These control schemes can be applied independently or collectively, thereby saving the energy spent on mechanical refrigeration by using airside economization and/or evaporative cooling. Various combinations of system-level controls and component-level controls are compared to a baseline no-controls case. The results show that reliability is consistently met by employing only sophisticated component-level controls. However, the recommended conditions are met approximately 50% of the simulated time by employing system-level controls only (i.e., SCADA) but with a reduction in data center cooling system power usage effectiveness (PUE) values from 3.76 to 1.42. Moreover, the recommended conditions are met at all averaged times with an even lower cooling system PUE of 1.13 by combining system-level controls only (SCADA and ON/OFF controls). Thus, the study introduces a simple method to compare control schemes for reliable and energy-efficient data center operation. The work also highlights a potential source of capital expenses and operating expenses savings for data center owners by switching from expensive built-in component-based controls to inexpensive, yet effective, system-based controls that can easily be imbedded into existing data center infrastructure systems management. [DOI: 10.1115/1.4044129]

Keywords: energy efficiency, data center, power usage effectiveness, controls

1 Introduction

U.S. data center energy consumption was 70×10^9 kWh in 2014 and is expected to increase to 73×10^9 kWh by 2020 per a recent study by Lawrence Berkeley National Laboratory (LBNL) [1]. The primary and secondary sources of power consumption in a data center are the information technology (IT) equipment and heating, ventilating, and air conditioning (HVAC) equipment, respectively. In recent times, however, these two areas of power consumption have become comparable. Many techniques to improve data center energy consumption rely on experimentally guided computational fluid dynamics modeling [2], computational fluid dynamics-like numerical modeling [3], and thermodynamic modeling [4].

The standard metric to gauge data center cooling system energy efficiency is the cooling system power usage effectiveness (PUE), defined as

$$PUE = \frac{cooling\ load + ITE\ load}{ITE\ load} \qquad (1)$$

Cooling load directly impacts cooling energy consumption, so it also has a direct impact on the cooling system PUE. Conventional (noncontrol) strategies to reduce the cooling system PUE include techniques such as cold and hot aisle containment, which reduce the recirculation of hot exhaust air from the IT equipment, thereby decreasing cooling airflow flow requirements and hence the cooling system power consumption. Cold aisle containment, however, leads to lower IT equipment inlet air temperatures [5]. Air-side economization, water-side economization, evaporative cooling, and warm water cooling are other commonly used strategies.

Relying solely on outside air rather than mechanical cooling can save up to 74% of cooling energy [6]. Similarly, water-side economization with warm water cooling for water-side chillers can save up to 17% energy [7], while direct evaporative cooling (DEC) has the potential to reduce data center energy consumption by 9% after taking maintenance and water consumption into account [8].

Recently, other possible techniques to reduce data center HVAC energy consumption have been based on control strategies tailored to the operation of a data center. These strategies can be categorized according to several types, such as thermal, electrical, and computational. Using modern algorithms, data center controllers designed using computational strategies such as coupled IT and cooling loads [9] or by electrical modifications such as direct current microgrids [10] can lead to additional power savings. A thorough review of various data center thermal management strategies yields only a select number of studies that explore the influence of different system-level thermal control strategies on PUE and reliability concurrently [11]. Therefore, this study intends to isolate the effects of controls on both energy efficiency and reliability.

Most studies on the thermal aspects of data center controls focus on employing active controls versus static controls for cooling equipment such as chillers and computer room air conditioning (CRAC) units. Additionally, replacing single speed drives with variable frequency drives in CRAC units can reportedly save up to 65% in power [12], while a temperature adaptive control strategy can save up to 17% power in a water-side chiller [7]. Other studies have focused on power failure scenarios, and they recommend keeping cooling equipment on backup power using a generator or universal power supply (UPS) (for high-density facilities) with thermal storage devices to decrease restart time [13]. Chen et al. [14] suggested using predictive, as opposed to reactive, controls for thermal and energy control in data centers. They reported 34% and 30% energy savings using their predictive controller as compared to static and reactive controls, respectively.

---

[1]Corresponding author.

The thermal energy "supply chain" in a data center is an interesting concept explored by Walsh et al. [15]. They suggest using a thermal management philosophy based on the chip temperature (i.e., the source of heat). Their study suggests that merely increasing rack inlet temperatures may not lead to an increase in the data center grand coefficient of performance, $(\mathrm{COP_{Grand}})$, defined as the ratio of total power to cooling power, due to an increase in fan power at the system level. An increase in heat sink temperature instead leads to savings in cooling costs. Moreover, they claim that a reduction in heat sink resistance by 50% improves data center cooling efficiency by about 40% since less cooling power is required to achieve the same performance. In addition, exergy-based analysis has pointed to the CRAC unit as a major source of irreversibility in a data center, and an optimum operating point that minimizes irreversibility in the CRAC unit based on the volumetric flow rate of incoming air has been suggested [16].

Recent cooling system controls work focuses on implementing local controls to improve energy efficiency. Mohsenian et al. [17] applied remotely operated variable airflow panels in a contained cold aisle to provision air to racks with low levels of overcompensation. VanGilder et al. [18] showed that applying louvered dampers between a contained hot aisle and return plenum can allow for precise tuning of the airflow rate. Finally, Baxendale et al. [19] demonstrated that a simple proportional–integral controller implemented in MATLAB can be used to maintain a server inlet air temperature by modulating the CRAC return temperature.

However, none of the literature reviewed so far focuses on the simultaneous impact of system and component-level controls on PUE and reliability. Therefore, this work contributes to existing research in the data center field by predicting the relative impact of system-level controls and component-level controls on the data center cooling system PUE and server reliability. The study aims for reliable and energy-efficient data center operation by developing novel thermal control techniques applied to traditional data center cooling equipment.

This study is based on an in-house holistic thermodynamic flow network modeling (FNM) tool developed to model data center IT equipment and cooling systems using both the first and second law approach [20]. The study applies coarse optimization using the FNM tool for data center cooling system design optimization and response. Combinations of two unique control schemes supervisory control and data acquisition (SCADA and ON/OFF controls) have been assessed for a generic cooling system utilizing airside and waterside economization as well as mechanical cooling through a CRAC unit.

2 Modeling

2.1 Comparison of Energy Efficiency and Reliability. The impact of using various combinations of controls to modulate parameters that affect data center cooling energy consumption and IT equipment reliability is investigated in this study using the FNM tool introduced previously. The goal is to enable data center operators to achieve the lowest cooling system energy consumption possible while maximizing ITE reliability. Reliability in this study is defined as the percentage of total uptime that the IT equipment inlet air conditions are kept in the ASHRAE recommended range (as opposed to allowable ranges) [21]. Traditionally, data center owners have been inclined to sacrifice energy efficiency to maintain reliability, which is seen in most overcooled facilities.

Table 1 illustrates how reliability is related to the ASHRAE server inlet condition ranges: recommended and allowable A1 through A4. The recommended range provides the largest reliability compared to the other ranges, with A4 providing the lowest reliability of the group. However, the recommended range also contains the lowest flexibility, theoretically meaning that the minimum energy efficiency in this range is higher than that associated with the other ranges, culminating with the highest energy

efficiency in range A4. This work allows data center operators to quantify the time spent in each ASHRAE range depending on choice of cooling system used, allowing users to strike a balance between energy efficiency and reliability, thus maximizing profits while lowering capital expenditure on infrastructure.

2.2 Virtual Test Bed. The cooling system modeled for this study is shown in Fig. 1. This typical system consists of airside economization aided by an evaporative cooler. Outside air is drawn in to cool the IT equipment when ambient conditions are conducive, such as low external temperatures and relative humidity, which are commonly prevalent in the winter months. However, during summers, ambient conditions prohibit the use of free air cooling, so data center operators often have to rely on mechanical cooling to cool the IT equipment. Such cooling systems commonly contain standalone DX-based CRAC units or CRAH units using air-cooled or water-cooled chillers, which further utilize cooling towers. Since both chillers and DX-based CRAC units are power intensive, this study uses the latter to reduce the number of cooling components and hence to focus on control schemes implementation rather than extensive cooling components modeling.

The virtual test bed shown in Fig. 2 has been developed to mimic the chosen data center cooling system and to test out the control schemes. It consists of an airside economizer to simulate free air and evaporative cooling, and a CRAC unit for mechanical cooling. Outside air is drawn into the mixing junction J2 using the supply air (S/A) fan, while the same amount of air is exhausted from the data center. Mixed air is then fed into the CRAC unit from where it is supplied to the data center airspace. The use of outside air can help in reducing CRAC runtime, thereby reducing data center operating cost.

Humid air and water were the two fluids used in the FNM. State air properties were calculated using a built-in psychrometric calculator, whereas water enthalpy was calculated assuming a fixed specific heat capacity of 4179 J/kg K. Further information on thermal-fluid calculations and properties can be found in Ref. [20].

2.3 Components' Model Description. Three key component models are used in this virtual test bed: an evaporative cooler, a direct expansion (DX)-based CRAC unit, and the data center airspace. In addition, models for simulating fans/pumps, dampers, and flow junctions are also used. Junctions are modeled as massless nodes with zero volume that assume perfect mixing of the inlet fluid streams and uniform properties among all outlet streams. A simple mass averaging of inlet flows for converging junctions is performed in the junction to calculate the exit flow enthalpy

$$\sum_i \dot{m}_i h_i = h_o \sum_i \dot{m}_i \tag{2}$$

where $\dot{m}$ is mass flow rate. A split flow maintains the inlet temperature for diverging flows.

Evaporative coolers are modeled as a medium with an associated single efficiency metric that is applied with the assumption of adiabatic cooling and full mixing of the air and water streams. The cooler's efficiency, η, is considered static and is defined as

$$\eta = \frac{T_{\mathrm{ao}} - T_{\mathrm{ai}}}{T_{\mathrm{wb}_i} - T_{\mathrm{ai}}} \tag{3}$$

where T_{ai} and T_{ao} are the inlet and outlet air dry-bulb temperatures, respectively, and T_{wb_i} is the inlet air wet bulb temperature. One can see that a "perfect" cooler would have air exiting at the wet-bulb temperature. A reasonable efficiency of 90% [22] is assumed for the current study. No change in temperature for the water stream is assumed.

Two different CRAC unit models (basic and advanced) are used in this study in both steady-state and transient modes. Both

Table 1 ASHRAE datacom equipment operating envelopes and their associated relative reliability and energy efficiency [22]

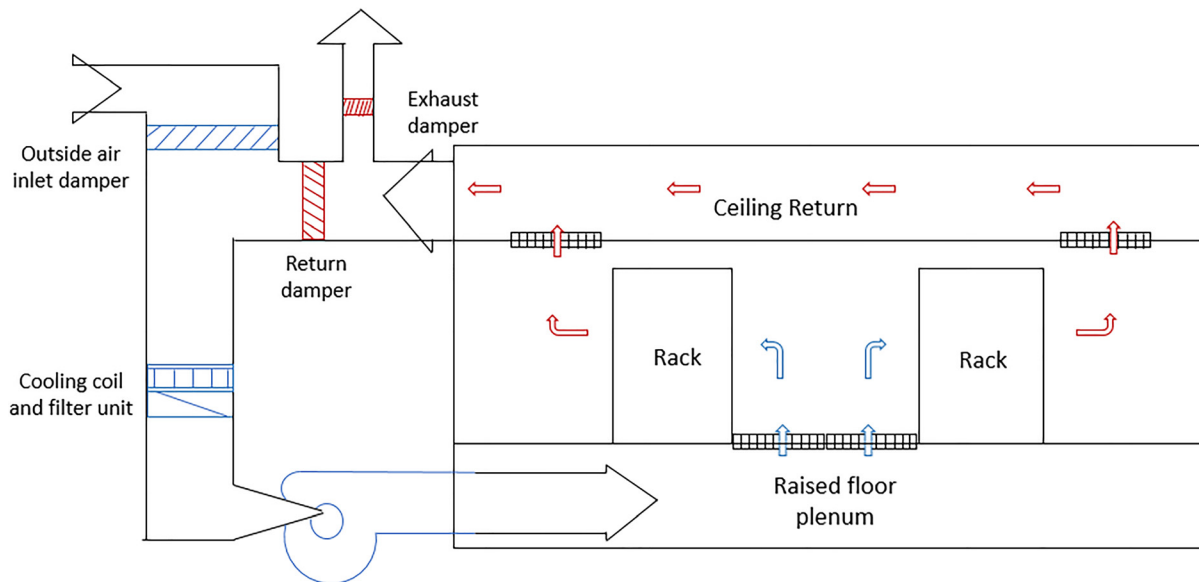| Range	Class	Dry-bulb temperature (°C)	Humidity range, noncondensing (%)	Maximum dew point (°C)	Theoretical reliability/ peak energy efficiency
Recommended	All A	18–27	5.5 °C DP to 60% RH and 15 °C DP	15	Very high/very low
Allowable	A1	15–32	20–80% RH	17	High/low
	A2	10–35	20–80% RH	21	Mid. range/mid. range
	A3	5–40	−12 °C DP and 8% RH to 85% RH	24	Low/high
	A4	5–45	−12 °C DP and 8% RH to 90% RH	24	Very low/very high

Fig. 1 Schematic of chosen data center cooling system. Adapted from Ref. [23].
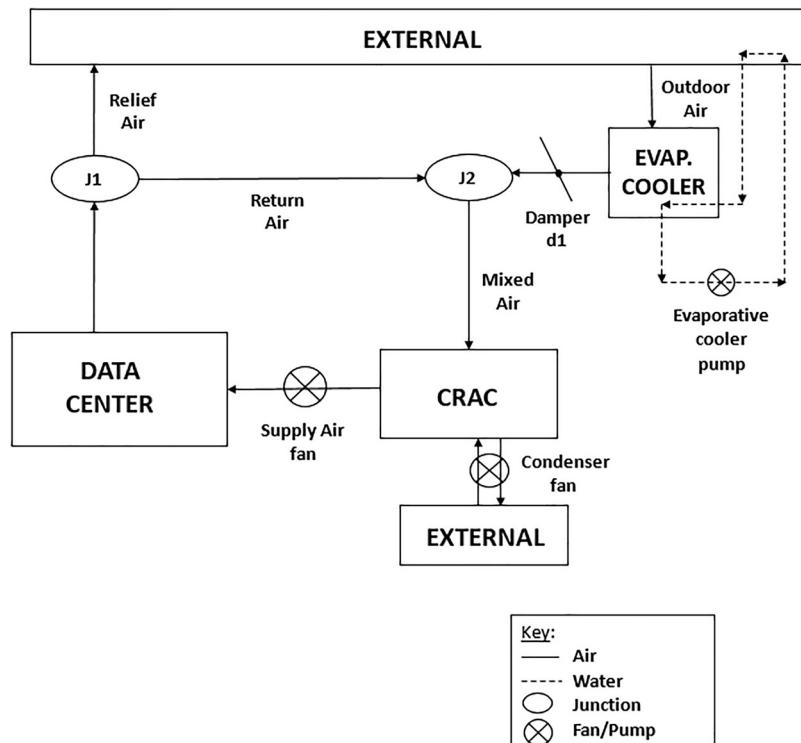

Fig. 2 Virtual controls test bed used in this study

levels feature analysis of all four stages of an ideal refrigeration cycle. The evaporator and condenser are assumed to have a heat exchanger thermal mass flow rate ratio of zero due to the dominance of refrigerant phase change over single-phase heat transfer and assume perfect heat exchange with the surrounding air. Thus, the standard $\varepsilon - NTU$ relation for a heat exchanger can be used [24]. Process 1–2–3 in Fig. 3 shows the operation of the basic CRAC model, while the advanced CRAC model can follow any of the three processes outlined on this figure.

The total heat load in the data center consists of server ($\dot{Q}_j$) and other electrical equipment ($\dot{Q}_e$) with no influence of external work. The airflow rate is assumed to be steady. The supply plenum, return plenum, and data center airspace each consist of a single volume of air with uniform temperature and humidity, while the airspace is conservatively modeled as a single spatially uniform temperature, with no distinction of hot and cold aisles. However, more advanced airspace models do exist in our FNM tool [25].

This simple model holds in cases when an average PUE value is of interest and the control schemes are based on supply air instead of rack inlet temperature. However, the model fails to capture data center transients, particularly air infiltration between the hot and cold aisle, recirculation of hot air in the rack, variation in inlet air temperature with vertical distance from the floor tiles, and when containment strategies such as hot aisle/cold aisle containment are used.

The three fans/pumps used here have known typical flow rates. However, the flow rate for the CRAC S/A fan is modulated by SCADA controls based on data center cooling requirements. An isentropic transition is assumed for the fans/pumps and the fluid (air or water) is considered incompressible. The pressure rise ΔP

for each fan/pump is calculated, allowing for an estimate of fan/pump power draw using

$$\dot{W}_{\mathrm{fan}} = \frac{Q_{\mathrm{fan}} * \Delta P}{\eta_{\mathrm{fan}}} \tag{4}$$

Similarly, an opposed blade action damper with a characteristic ratio of 10% is used for this study [26] and its position is varied using SCADA control to minimize the data center's reliance on mechanical cooling. Dampers are modeled as flow control devices with an approximated flow coefficient (C_v) profile [27]. Manufacturer-provided velocity and pressure values for a pipe with a given hydraulic diameter are converted to C_v values using [28], using which the dimensionless loss coefficient K_L can be determined as

$$K_L = 891d^4/C_v^2 \tag{5}$$

where d is the hydraulic diameter of the duct that contains the damper, and the proportionality constant is empirically determined. Finally, knowing the characteristic profiles of parallel and opposed blade dampers [26], a loss coefficient profile for the damper can be determined for various damper openings. An equal percentage curve has been used in our model for this purpose.

2.4 Simulation Parameters. Dallas-Fort Worth, TX, was chosen as the simulation location due to its hot and humid weather during the summer months. This provides a challenging location to test the efficacy of the control schemes as compared to locations with ambient conditions more favorable for free air cooling.
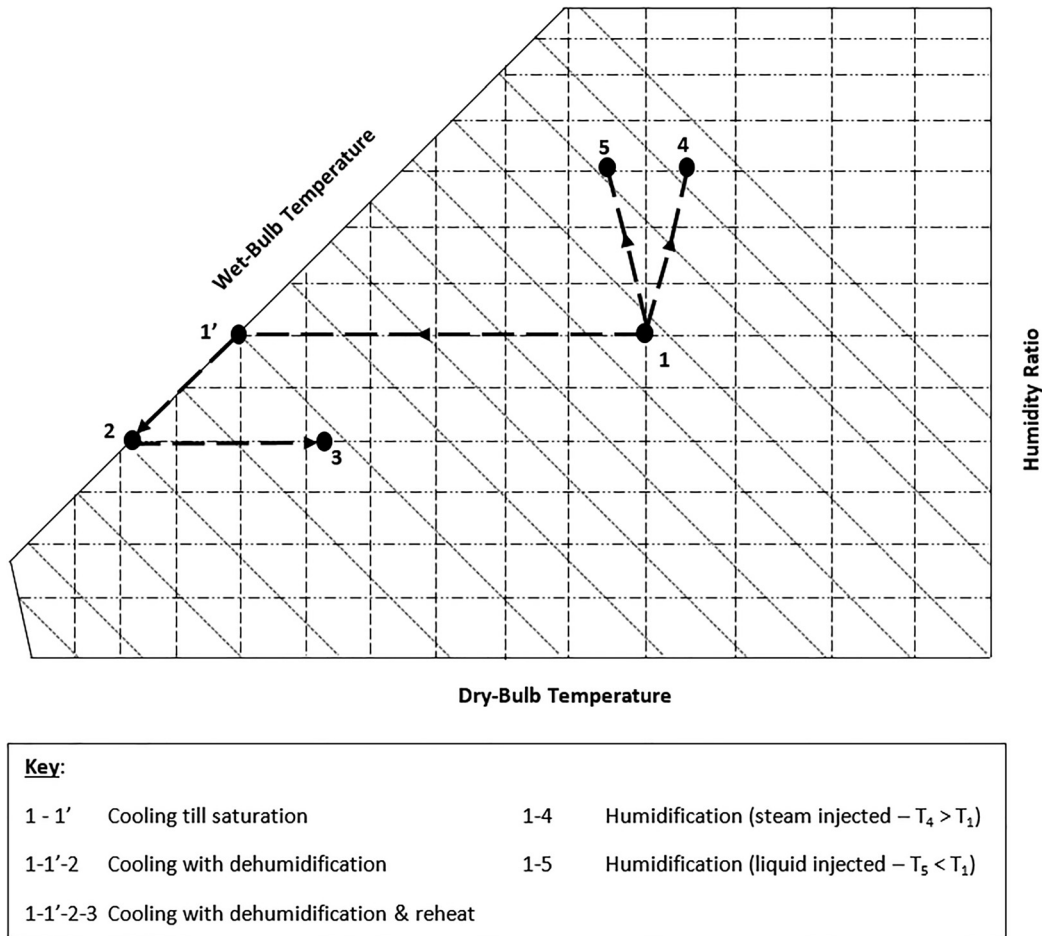

Key:

1 - 1'	Cooling till saturation
1-1'-2	Cooling with dehumidification
1-1'-2-3	Cooling with dehumidification & reheat

1-4	Humidification (steam injected – $T_4 > T_1$)
1-5	Humidification (liquid injected – $T_5 < T_1$)

Fig. 3 Psychrometric chart showing working methodology for both CRAC models

Simulations were run for four days in July, from the 20th to the 23rd. The first three days were used for proper initialization of the controls, while results from the 23rd were kept since that date contained the highest annual enthalpy per an NREL database [29].

The time-step was initialized as 10 s but allowed to increase to a maximum of 120 s by scaling at each step when the change in a chosen state variable value is below 10% from the previous step per the metric

$$\varepsilon_{\text{curr}} = \left| \frac{SV_{\text{curr}} - SV_{\text{prev}}}{SV_{\text{curr}}} \right| \qquad (6)$$

where SV represents the state variable, chosen to be the lumped airspace temperature in this case, and a constant scaling factor of 2 is used throughout all simulations. In this manner, the timestep is slowly increased to make the simulations run faster while ensuring stability of results. However, if the current error is larger than the specified threshold of 10%, then the next timestep is decreased.

An implicit third-order backward difference scheme was used to solve the two sets of equations in matrix form ($A\mathbf{x} = b$) for absolute humidity and enthalpy for each lump in the system. A third-order scheme provides stability for relatively large time steps in excess of 60 s, which makes the simulation computationally efficient as opposed to a first-order scheme that quickly becomes unstable at the time steps employed here.

A tolerance of 10^{-7} to be achieved in 1500 iterations was used as the convergence criteria for the flow solver, which solves for the path flow resistances and flow rates using equations (B1) and (B2). A relaxation factor of 0.01 was employed to aid in convergence of the flow solver. In addition, a combined tolerance of 10^{-3} with a maximum of 40,000 iterations at each timestep was used as the convergence criteria for the energy and mass balance solvers. For a timestep to successfully converge, the flow solver must first successfully converge. Following this, the error from both the energy and mass solvers at each timestep is accumulated until it drops below the specified tolerance within the maximum number of iterations specified. The simulation then proceeds to the next timestep and solves it in the same manner until the specified simulation end time is reached.

Simulations featuring the same conditions but without controls were run for comparison. The parameter values for the no-controls case were determined from the SCADA database as those corresponding to the maximum enthalpy case and are shown in Table 2.

Inspection of Table 2 reveals that the optimum values of the three parameters of interest are within 10% of each other, confirming the lack of influence of any external controls on the overall system. However, the deviation exhibited by the CRAC S/A temperature is due to the presence of internal controls in the advanced CRAC model, which lowers the optimal value of the setpoint. This invariably leads to greater power consumption by the advanced CRAC model, which is expected due to its ability to meet the temperature and humidity setpoint at each timestep.

3 Control Strategies

3.1 System Dynamics and Controller Architecture. In the virtual test bed of Fig. 2, the damper position directs the amount of supply air into the data center from outside. Mixing at junction

Table 2 Optimized values for CRAC fan flow rate, damper position, and CRAC supply air setpoint for the no-controls case

	Fan f3 flow rate (m³/s)	Damper d1 position (deg)	CRAC S/A setpoint (°C)
CRAC basic	0.369	4.51	19.2
CRAC advanced	0.369	4.06	17.8

J2 dictates the conditions of the air being fed into the CRAC unit, and—depending on its setpoint—further determines the level of CRAC mechanical cooling. Minimizing CRAC runtime can lead to energy savings and a reduction in data center cooling system PUE values. Thus, key parameters are identified as those that most affect cooling system dynamics; in this case, the CRAC fan (f3) flow rate, CRAC S/A temperature, and damper (d1) position are modulated by SCADA controls.

3.2 Supervisory Control and Data Acquisition Control. Figures 4(a) and 4(b) show a general controller architecture developed for this study. Figure 4(a) focuses on the development of the SCADA database, whereas Fig. 4(b) demonstrates the application of control schemes in a transient simulation. SCADA control is based on a database of optimum values for the key parameters for different external conditions.

Creating the database requires modulation of these parameters using Monte Carlo sampling. The steps for creating a SCADA database are shown as a flowchart in Fig. 5, following which three parameters were identified as key to data center energy efficiency in this test bed. These include the CRAC fan (f3) flow rate, CRAC S/A temperature, and damper (d1) position. The CRAC unit's compressor is the principal source of energy consumption, while the damper can allow for airside economization, thus reducing the operating time for the CRAC unit and hence saving on cooling system energy consumption, thus enhancing the PUE value of the data center.

The ASHRAE recommended range of temperature and humidity were used as constraints for server reliability. Specifically, the dry-bulb temperature and dew point must be between 18 °C and 27 °C and between 5.5 °C and 15 °C, respectively, and the relative humidity is less than 60% to make it five constraints. The cooling system PUE is chosen as the optimization criteria.

For this study, the external conditions are varied from a low value to a high value (i.e., the external temperature ranges from 10 °C to 40 °C, and the external relative humidity varied from a relatively dry value of 20% to an extremely humid value of 100%) in order to cater to most climates within which a data center can be located. This extensive variation calls for different ranges of parameters to be used for each set of conditions. In each case, the range is set to balance (1) the availability of viable data points that meet the constraints, and (2) the resolution of optimal cooling system PUE values. For example, for a high ambient temperature of 40 °C, the damper position would be kept to a minimum to prevent hot outside air from entering the data center. However, an ambient temperature of 20 °C would readily allow for airside economization, and hence, the damper opening can be far greater than the previous case.

Furthermore, 100 random samples were chosen for each set of external conditions to balance the computational time and accuracy of the optimized solution set. Trial runs with sample sizes of 500 yielded results within 10% of those with the smaller set. However, the SCADA database creation time increased by nearly a factor of three.

Appendix C and Fig. 11 visually represent the SCADA database generation process. A database of optimal values for the three key parameters was thus created for the range of external conditions given previously. At each time-step of a simulation, based on external conditions, the SCADA controller calculates the best value for each parameter using 2D interpolation on the database.

3.3 ON/OFF Control. ON/OFF controls switch "on" or "off" specific components. In the OFF state, a cooling component does not draw any power, which reduces power consumption. With this goal in mind, ON/OFF controls are used to control the state of the CRAC unit since it is the most significant power consuming component in the cooling system. It is important to note that this system-level ON/OFF control is separate from the component
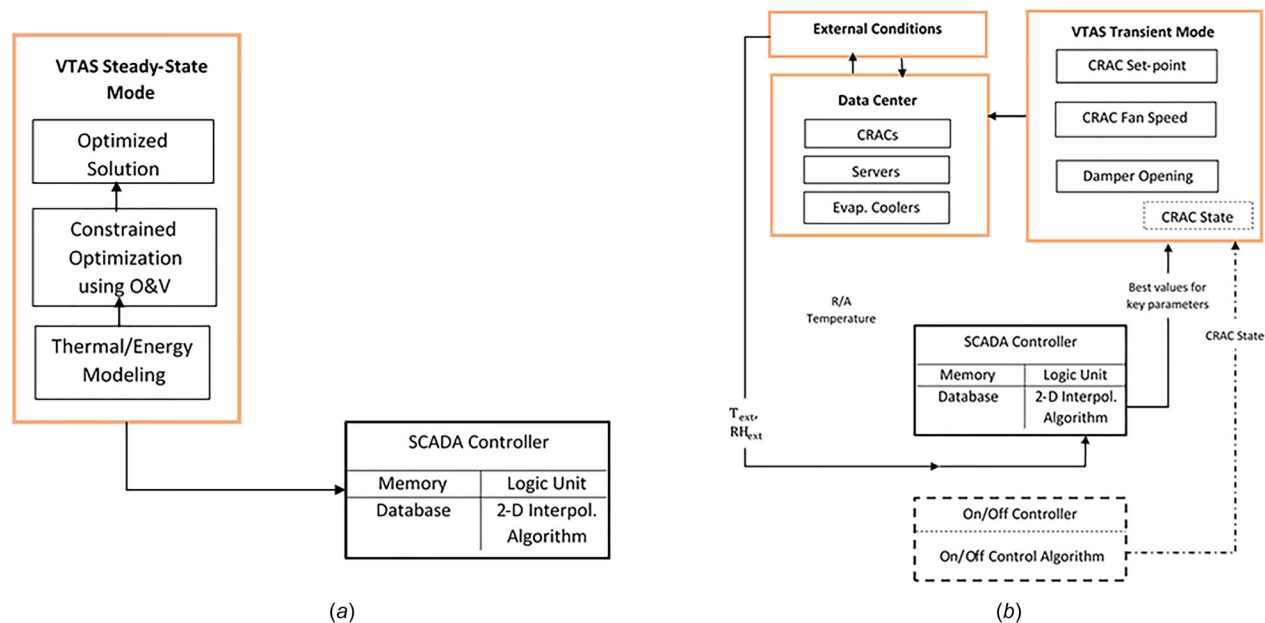
Fig. 4 (*a*) VTAS controller architecture in steady-state mode and (*b*) VTAS controller architecture in transient mode

built-in ON/OFF control found in the advanced CRAC model. The system level ON/OFF controller functions using the logic described in the flowchart of Fig. 6. A deadband (db) value of $\pm 0.5 \,^{\circ}$C was used for all simulations.

The use of a deadband prevents rapid ON/OFF cycling of the CRAC unit, thus ensuring compressor reliability. It should be noted that in this case, the data center power density is 0.845 kW/rack, which is roughly five times less than the average power density for most modern data centers (3–5 kW/rack) [30].

Fig. 5 Flowchart for SCADA database creation

Moreover, for the combined SCADA and ON/OFF controls (case e), which yields the lowest PUE and the highest reliability, the CRAC unit switched on and off six times during a 24-hour cycle on the maximum enthalpy day. Hence, the frequency with which the CRAC unit switches ON or OFF is not seen as an issue if this controller is implemented in a real data center.

It should further be pointed out that this rudimentary control was implemented to augment the advanced SCADA controls and primarily not intended to function on their own. However, if implemented in a high-density data center, frequent ON/OFF switching of the CRAC unit can be a problem leading to poor compressor reliability. In that case, either the CRAC setpoint can be lowered enough so as to prevent frequent switching or the deadband increased by the CRAC manufacturer for these high-density applications.

For the advanced CRAC model, the ON/OFF controller functions the same way as its system-level counterpart and employs the same deadband values, thus allowing for an equal comparison between the two control schemes.
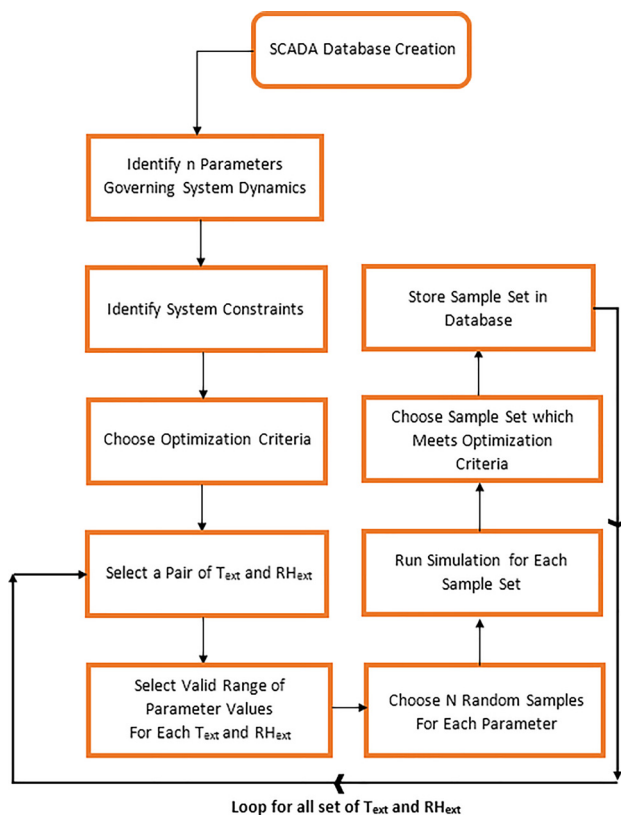
3.4 Cooling System Power Usage Effectiveness Calculation. The cooling system PUE is determined using Eq. (1). For the airspace model used in this study, our model does not account for server fan load, and the calculation of ITE load only includes the server heat load. The ITE load used for this study was 2535 W while the cooling load varied depending on external conditions and the choice of controller. A separate investigation on a similar data center using a higher order airspace model revealed that the server fan load only totaled to around 5 W and thus has a negligible effect on cooling system PUE. Thus, the server fan load was ignored in this study.

4 Results and Discussion

The virtual test bed shown in Fig. 2 was employed for three different cases (no control, ON/OFF control and SCADA control). The SCADA database of key parameter optimum values along with the corresponding cooling system PUE are shown in Tables 3 and 4 for the basic and advanced CRAC unit models, respectively. The tables show that the choice of CRAC model plays a major factor on the optimal parameter settings. Furthermore, traversing down Tables 3 and 4 suggest that external temperature has more influence on optimal parameter settings compared to the external
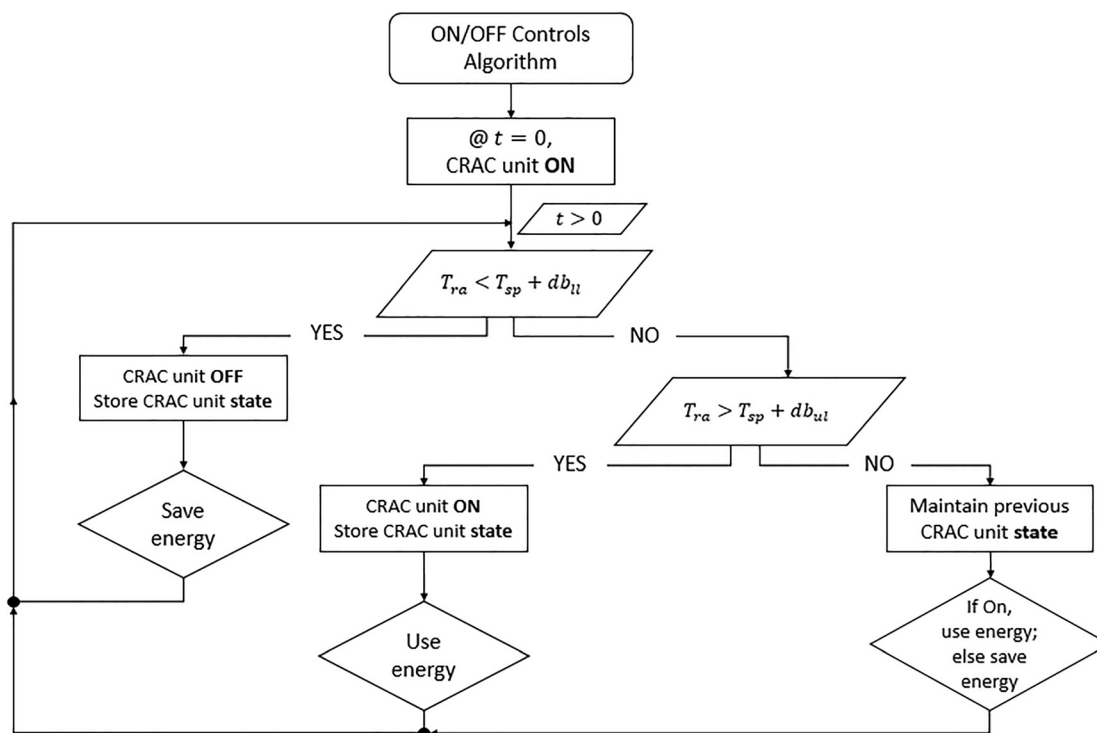
Fig. 6 ON/OFF controller logic. "db" represents controller deadband.

relative humidity, as evidenced by the significant change in parameter values going from block to block versus a slight perturbation within a block of fixed external temperature. Furthermore, the lowest PUE values are obtained for the minimum external temperature considered (i.e., $10\,°C$) as this allows the supply air temperature setpoint to be attained more often, thus saving on the power required to run the CRAC compressor. A holistic comparison of Tables 3 and 4 reveals higher PUE values in general for a given set of external conditions for the advanced CRAC model versus the basic model. This can be attributed to the advanced CRAC's ability to always meet the temperature and humidity setpoint, which may require cooling below the dew point and subsequent reheat, both which consume more electric power as opposed to the simple cooling mechanism of the basic CRAC model.

All simulations were run on a Dell OptiPlex 7040 personal computer with a 2.0 GHz Intel Core i5 processer and 8 GB of RAM. Simulation times varied with the longest simulation featuring SCADA controls using the advanced CRAC model taking about 37 min to run, while the shortest simulation employing no-controls and the basic CRAC model took only 15 min to complete. Tolerances of 10^{-7} and 10^{-3} were used as the convergence criteria for the flow solver and energy and mass solvers, respectively, at each timestep. Further details regarding simulation parameters are specified in Sec. 2.4.

Computer room air conditioning units were sized using maximum ambient enthalpy as the worst-case scenario (July 23rd at 3 pm for Dallas-Fort Worth). While the CRAC unit supply air temperature setpoint was varied, its supply air (S/A) humidity setpoint was fixed at 45% throughout to ensure compliance with the ASHRAE recommended inlet air relative humidity of less than or equal to 60%. All simulations were run from July 20th through 23rd. Results from the first three days were used to initialize the controls and hence discarded, and only results from the fourth day were used for analysis. Other parameters remained the same throughout all simulations.

Figure 7 shows the variation of external temperature and relative humidity with time for the maximum ambient enthalpy day (July 23rd).

Each horizontal bar on the temperature curve represents an external temperature value obtained by averaging all temperature data points in an hour's interval (e.g., the first point represents a temperature value obtained by averaging data between midnight and 1 am on July 23rd), up to midnight the following day, where the last point represents data averaged between 11 pm on July 23rd and midnight of the next day. Similarly, the diamond marker curve represents averaged external relative humidity data. The figure shows peak temperatures reaching nearly $36\,°C$ and relative humidity values up to 79%, confirming that the climate is hot and humid.

Table 3 Optimized values for CRAC fan flow rate, damper position and CRAC supply air setpoint using the basic CRAC model

Temperature (°C)	Relative humidity (%)	Fan f3 flow rate (m³/s)	Damper d1 position (deg)	CRAC S/A setpoint (°C)	Cooling system PUE
10	20	0.644	22.8	22.8	1.14
10	40	0.695	24.4	22.8	1.13
10	60	0.659	23.1	22.9	1.21
10	80	0.698	24.8	22.9	1.22
10	100	0.699	24.4	22.9	1.27
20	20	0.595	24.4	21.7	1.31
20	40	0.693	21.6	21.9	1.40
20	60	0.453	24.8	21.9	1.47
20	80	0.361	10.2	20.0	1.51
20	100	0.356	11.4	20.0	1.53
30	20	0.401	5.03	20.0	1.50
30	40	0.412	6.43	20.0	1.52
30	60	0.412	5.56	19.9	1.53
30	80	0.411	6.17	20.0	1.56
30	100	0.408	5.52	19.9	1.57
40	20	0.317	2.76	18.0	1.51
40	40	0.304	3.32	18.0	1.53
40	60	0.301	2.45	17.9	1.53
40	80	0.308	2.86	18.0	1.55
40	100	0.302	2.58	17.9	1.56

Table 4 Optimized values for CRAC fan flow rate, damper position, and CRAC supply air setpoint using the advanced CRAC model

Temperature (°C)	Relative humidity (%)	Fan f3 flow rate (m³/s)	Damper d1 position (deg)	CRAC S/A setpoint (°C)	Cooling system PUE
10	20	0.320	11.6	18.0	1.32
10	40	0.302	10.2	18.9	1.29
10	60	0.700	23.4	18.4	1.22
10	80	0.696	24.1	20.6	1.21
10	100	0.596	24.8	22.5	1.20
20	20	0.594	29.7	20.9	1.21
20	40	0.301	24.8	19.2	4.20
20	60	0.352	10.1	19.0	4.71
20	80	0.352	10.5	20.0	4.73
20	100	0.353	10.5	20.0	4.76
30	20	0.408	5.33	17.5	5.23
30	40	0.401	5.08	17.5	5.19
30	60	0.412	5.31	17.7	5.30
30	80	0.400	5.03	17.9	5.21
30	100	0.402	5.03	18.2	5.25
40	20	0.300	2.22	17.9	4.22
40	40	0.301	2.08	18.0	4.24
40	60	0.305	2.11	17.9	4.29
40	80	0.313	2.57	18.0	4.39
40	100	0.308	2.56	18.0	4.36

Data center PUE values and airspace dry-bulb and dew point temperatures were predicted using five different control schemes: no control (case a), SCADA controls only (case b), CRAC controls only (case c), SCADA and CRAC controls (case d), and SCADA with ON/OFF control (case e). Cases a, b, and e use the basic CRAC model, whereas cases c and d use the advanced CRAC model. These are represented in Fig. 8.

Figure 8 shows that having no controls at all (case a) leads to poor reliability of the IT equipment, as evidenced by the stated ASHRAE range for mission critical facilities [21]. Adding sophisticated component level controls (case c) alone, such as those found in a typical CRAC unit, can easily maintain data center airspace conditions in the ASHRAE recommended range, which increases the reliability of the IT equipment. However, this approach comes at the cost of high equipment power consumption as indicated by the high cooling system PUE values of around 3.7, which would be unacceptable for a typical data center because of the wasted energy. The addition of system level controls to these component level controls (case d) only adds to the power consumption, further raising PUE values, while still maintaining reliability and hence does not promise to be economically viable from a data center owner's perspective.

On the other hand, employing sophisticated system level controls (case b) with a basic CRAC unit seems to provide relatively greater reliability than when no controls are used (i.e., case a) while reducing power consumption. The figure shows that using SCADA controls alone can keep the IT equipment in the ASHRAE recommended range for 50% of the time, while the no-controls case never achieves this range.

Figure 9 shows the average cooling system PUE to be nearly identical for the SCADA controls and no-controls case for this maximum enthalpy day. Thus, employing system level controls such as SCADA promises to maintain reliability longer while minimizing energy consumption. However, the issue with SCADA controls in this case is that they are not able to fully control the cooling system dynamics since there is no built-in temperature control for the basic CRAC unit. SCADA controls can only modulate the CRAC setpoint temperature; however, the basic CRAC unit model does not have the ability to meet that setpoint 100% of the time, thus leading to lower reliability of the air supplied to cool the IT equipment.

Finally, a hybrid controller (case e), which combines the sophistication of SCADA controls with the simplicity of ON/OFF controls, fully captures the system dynamics. Combining the switching capability of ON/OFF controls with SCADA controls allows it to switch the basic CRAC unit off when it overcools, hence preventing overcooling of the data center airspace and subsequently saving energy. Figures 8 and 9 compare this hybrid controller to the other control schemes presented. The figures show that case e maintains 100% reliability while keeping energy consumption to a minimum, as evidenced by the lowest cooling system PUE value (1.13) of the group and the ITE inlet air conditions being present in the ASHRAE recommended range for 100% of the simulated time. It should be noted that simulations were also performed for case (e) featuring system ON/OFF control without a deadband, and the calculated cooling system PUE values were identical to those shown in the table to three significant figures, thus confirming that it is indeed the combination of these two control schemes alone and not the influence of any other parameter, control-related or external, which drives the data center to achieve the best conditions of PUE and reliability among all the cases.

Finally, it is worth mentioning that the thermal mass of the servers influences their thermal characteristics leading to transients in the data center, which impacts the performance of the cooling system and ultimately the data center energy consumption. The server thermal mass can be characterized in terms of a
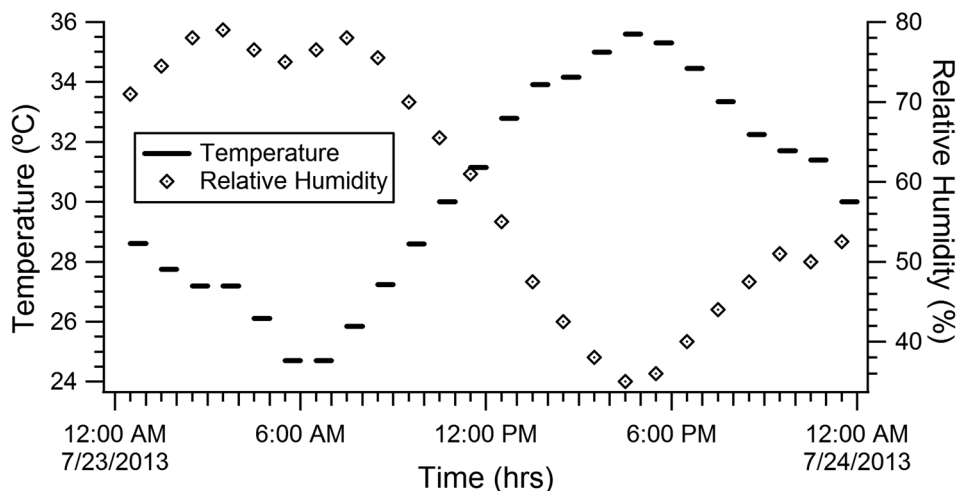
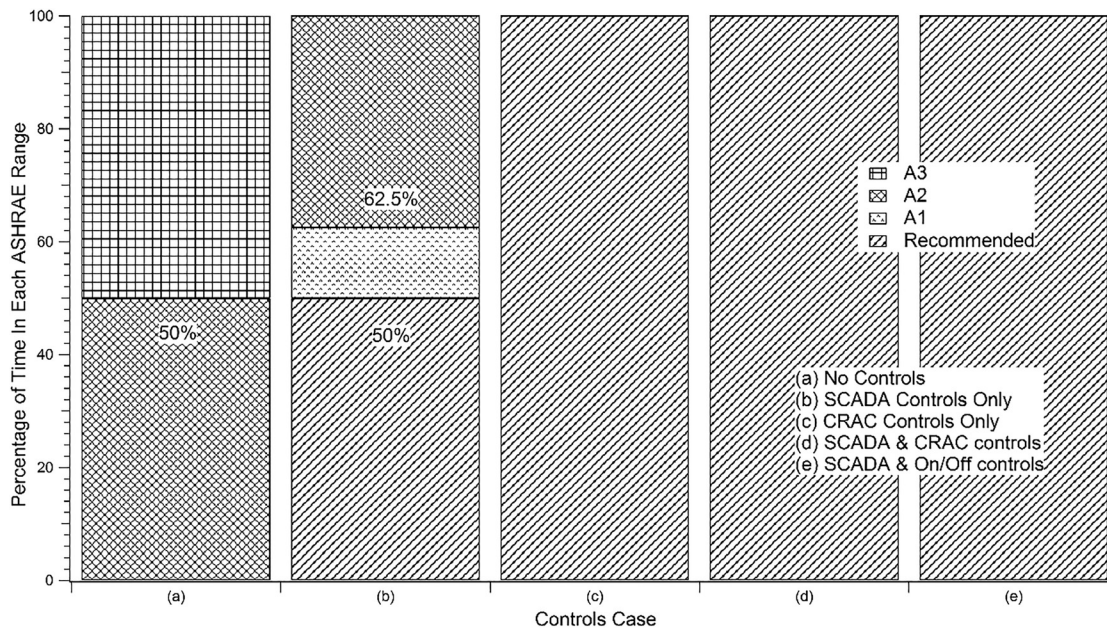**Fig. 7 Variation of external temperature and relative humidity for July 23rd in Dallas–Fort Worth**

Fig. 8 Comparison of time spent in each ASHRAE range for each control scheme
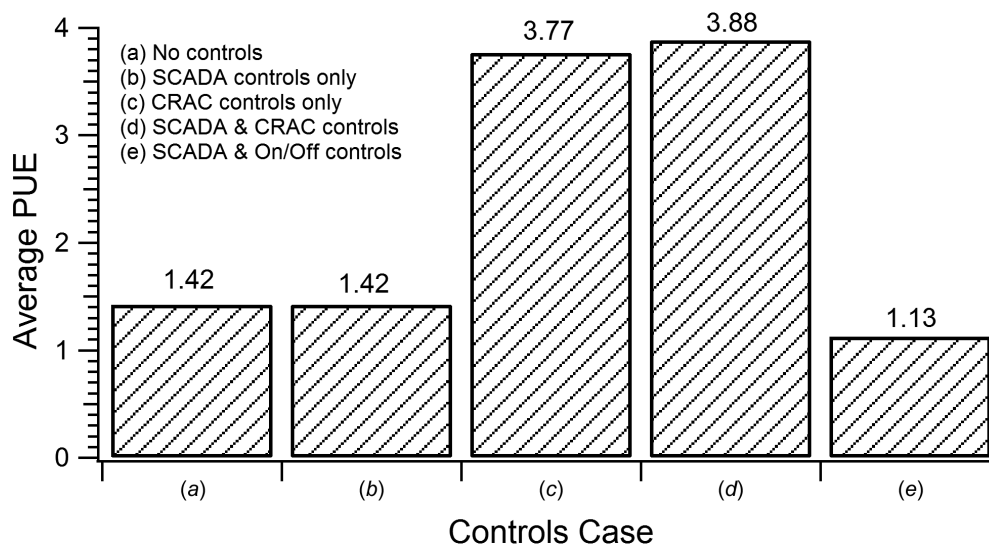

Fig. 9 Comparison of average cooling system PUE for each control scheme

thermal capacitance and time constant. The influence of server thermal mass was analyzed as a test case. Based on Refs. [31] and [32], an average thermal capacitance value of 11.1 kJ/K/server was chosen, and simulation cases (a) and (e) (no controls and hybrid SCADA and ON/OFF controls) were re-run with the updated server characteristics. The results showed negligible difference in the results for the two cases to three significant figures. The finding may be attributed to the fact that our timestep values (10–120 s) is smaller than the server time constant, which from Refs. [31,32] is about 6–8 min, thereby effectively capturing transient behavior in the CRAC ON/OFF controls.

5 Conclusions and Future Work

This paper provides insight into the development work of two different types of system level control schemes. A virtual data center test bed with a lumped airspace model is utilized for use inside an in-house developed FNM tool. The lumped model suffices for purposes of control schemes' comparison while avoiding the detailed modeling of the thermal/fluid characteristics of the data center airspace.

Data center controls at the system level have been compared separately and in conjunction with those existing at the component level. For hot and humid days, simulation results suggest that using a combination of the developed system level controls alone (case e) may be adequate in ensuring continuous IT equipment reliability while keeping power consumption minimal as opposed to using component level controls alone (case c) or a combination of system level and component level controls (case d). Hence, incorporating the suggested system level controls into existing data center infrastructure systems management can potentially keep data center operating expenses and capital expenses to a minimum by saving on utility cost and the cost required to procure expensive cooling equipment, respectively.

Future work entails testing the suggested control scheme in a real data center to validate the simulation results. Additional

measures such as the use of local temperature sensors, in addition to IT sensors, and varying thermal loads should be utilized to determine the robustness of the suggested control scheme. Finally, server loads are transient in nature—both in terms of random and fixed changes—and these effects of transient server load magnitude and period on the energy efficiency and reliability of various controls schemes need to be addressed in a future study.

Acknowledgment

Funding Data

Nomenclature

C_v = flow coefficient (US gpm at 60 °F for 1 psi pressure drop across valve)
d = hydraulic diameter, m
db = deadband
h = specific enthalpy, J/kg
K_L = dimensionless loss coefficient
$\dot{m}$ = mass flow rate, kg/s
P = pressure, Pa
Q = volumetric flow rate, m^3/s
R = flow resistance, Pa·s^2/m^6
t = time, s
T = temperature, °C

Greek Symbols

ε = effectiveness
$\varepsilon_{\text{curr}}$ = convergence tolerance
η = efficiency

Subscripts

ai = air inlet
ao = air outlet
db = dry-bulb
dp = dew point
fan = fan
ll = lower limit
ma = mixed air
oa = outdoor air
ra = return air
sa = supply air
ul = upper limit
wb = wet-bulb

Appendix A: Brief Description of Flow Network Modeling With Villanova Thermodynamic Analysis of Systems

Villanova thermodynamic analysis of systems [20] predicts HVAC system flow rates based on user-defined quantities. In FNM, the calculations of flow rate and pressure drop involve the creation of lumps and paths. Lumps represent a volume of fluid with spatially uniform thermodynamic and psychrometric properties, while zero-volume paths connect the lumps, allowing fluid to flow between lumps. The general equation for pressure drop for fluid flow from a lump i to a lump j, derived in [33], is

$$P_i - P_j = R_{ij}Q_{ij}^2 - \Delta P_{\text{gain}} + \Delta P_{\text{loss}} \qquad \text{(A1)}$$

where R_{ij} is the flow resistance specified using the standard pipe energy equation, Q_{ij} is the volumetric flow rate, ΔP_{gain} is the net pressure rise due to a fan or pump, and ΔP_{loss} represents additional pressure losses not accounted for by the flow resistance. The flow resistance in connecting components that incur bends and turns in the flow as it moves from one node to the other is taken from Ref. [34]. A mass balance is applied on each lump except the final lump to complete the equation set. The above-mentioned nonlinear equations are then solved in an iterative fashion for values of P and Q^2, using a modified form of relaxation [35]. These flow rates are then used in creating an energy balance to determine the energy and exergy losses throughout the system in a steady-state configuration, which forms a starting point for transient simulations.

Appendix B: Villanova Thermodynamic Analysis of Systems Validation

Villanova thermodynamic analysis of systems is validated against experimental data provided by Panduit[2] from their data center laboratory using various containment strategies including hot aisle/cold aisle containment. Load banks of 319.7 kW are used to represent the IT load. Figure 10 shows the virtual schematic of their test facility used in the validation exercise.

The comparison of experimental and simulation results is shown in Table 5. The CRAH fan, primary pump, and secondary pump were modeled to have a specified flow with overall efficiency values of 40%, 80%, and 18%, respectively, calibrated to match experimental data. Since pump/fan efficiency varies parabolically with flow rate, it is possible to have very low efficiency values near the two flow extremes [36], as in the case of the CRAH fan and secondary pump. A K-factor of 2.65 was added to holistically simulate air flow characteristics and hence the pressure dropped through the racks and servers and to match the experimental CRAH fan power. Finally, an air-cooled chiller component model was developed using an artificial neural network (ANN) fit to 592 experimentally measured data sets. The model's predicted chiller power is off by 16% from the experimental value, which can be improved by using a larger data set.

Overall, there is a 10% discrepancy between experimental and theoretical results, largely due to the ANN-based chiller model. This discrepancy has the greatest impact on the resultant total data center power consumption but is reasonable enough to represent good agreement between the software and actual data center measurements.

Appendix C: Example of Monte Carlo Sampling Method

The process outlined in the SCADA control section is separately run for each pair of external conditions, and an example is shown in Fig. 11.
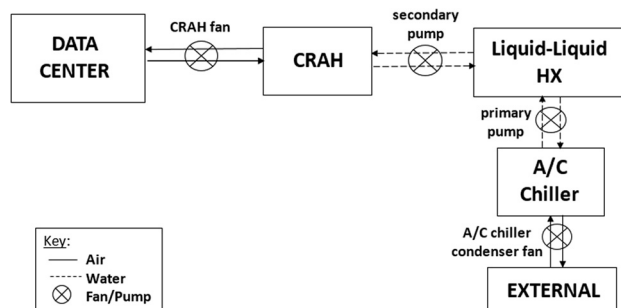

Fig. 10 Panduit test bed for VTAS validation

Table 5 Comparison of experimental versus VTAS simulation results using data from Panduit

	Experimental-maximum capacity	VTAS	Absolute error
IT load (kW)	319.7	319.7	User defined
CRAH			
Supply air temperature (°F)	63.10	63.10	User defined
Return air temperature (°F)	93.70	91.92	1.90%
Airflow rate (cfm)	35,874	35,874	User defined
Fan power (kW)	31.40	31.52	0.38%
Secondary chilled water			
Supply air temperature (°F)	44.85	44.85	User defined
Return air temperature (°F)	59.94	59.20	1.23%
Water flow rate (gpm)	152.6	152.6	User defined
Pump speed (%)	98.8	N/A	N/A
Pump power (kW)	4.97	4.83	2.82%
Primary chilled water			
Supply air temperature (°F)	39.34	39.34	User defined
Return air temperature (°F)	49.96	47.16	5.60%
Water flow rate (gpm)	280.0	280.0	User defined
Pump power (kW)	4.10	4.32	5.37%
Chiller			
Ambient air temperature (°F)	26.54	26.54	User defined
Compressor power (kW)	68.38	57.36	16.1%
Chiller utilization (%)	30	N/A	N/A
Condenser fan speed (%)	61.60	N/A	N/A
Overall cooling system			
Total power (kW)	108.9	98.03	9.98%
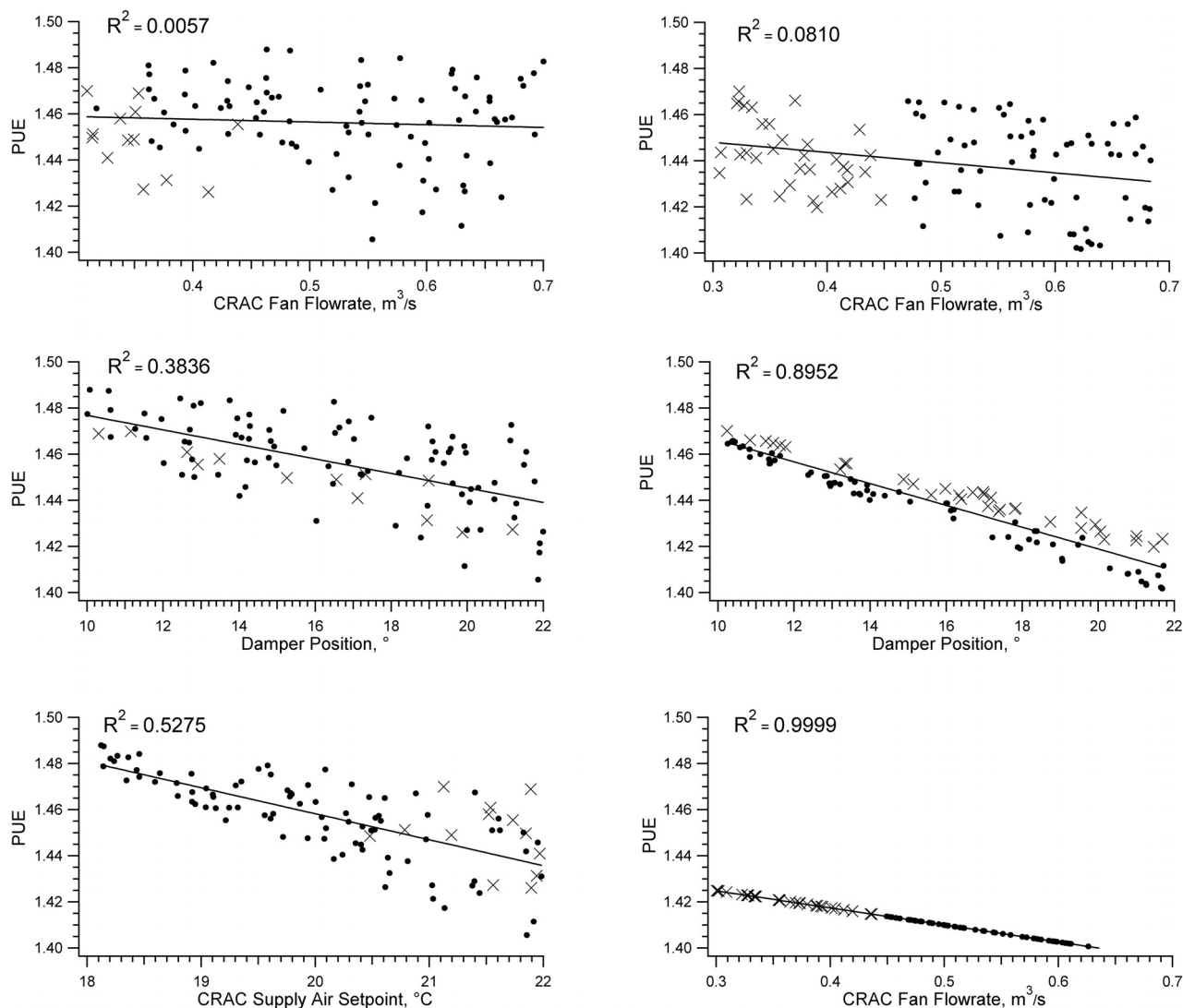
Fig. 11 Optimization of all three variables using regression and random sampling for external conditions of 20 °C and 40% relative humidity

In Fig. 11, dots represent samples that meet all the constraints while crosses represent samples that violate one or more of the constraints. The dots and crosses combined represent the total sample set for a given simulation run. Ambient temperature and relative humidity values of 20 °C and 40%, respectively, with a sample set of 100 values are used to illustrate the method.

In the left column of Fig. 11, the three control parameters of importance to this study are varied for the set of external condition stated previously. The figure shows three separate distributions, one for each parameter along with their respective correlation coefficient (R^2 value) and the line of best fit. This value dictates how well a given parameter is correlated with the overall system—the higher the number, the better the correlation. In this case, the CRAC S/A temperature contains the highest R^2 value, which is 0.528. Hence, the CRAC S/A temperature is considered optimized and a value of 21.85 °C is chosen, corresponding to the sample with the least cooling system PUE that also meets the constraint.

The process is next repeated with two variables with the CRAC setpoint fixed at 21.85 °C. Results are shown in the first two rows of the right column of Fig. 11. This time, the damper position has the higher correlation coefficient and is thus considered optimized. Hence, a value of 21.8 deg is chosen as one that meets the constraint and offers the least cooling system PUE.

Finally, the process is run one more time with the CRAC setpoint and damper position fixed at the above-given values. This optimizes the CRAC blower fan flow rate, and the results are shown in the third row of the second column of Fig. 11. The final variable being optimized yields a correlation coefficient value approximately equal to 1. A value of 0.695 m³/s is chosen as optimal for the CRAC setpoint, completing the parameter set for the given external conditions. The process is repeated for the remaining sets of external conditions. It should be noted that the subfigures in Fig. 11 show that the PUE is insensitive to the CRAC fan flow rate, indicating that natural modulations in flow rate even as high as ±20% will not significantly impact energy savings.

References

[1] Shehabi, A., Smith, S., Sartor, D., Brown, R., Herrlin, M., Koomey, J., Masanet, E., Horner, N., Azevedo, I., and Lintner, W., 2016, "United States Data Center Energy Usage Report," Lawrence Berkeley National Laboratory, Berkeley, CA, Report No. LBNL-1005775.

[2] Alissa, H. A., 2015, "Innovative Approaches of Experimentally Guided CFD Modeling for Data Centers," 31st Thermal Measurement, Modeling & Management Symposium (SEMI-THERM), San Jose, CA, Mar. 15–19, pp. 176–184.

[3] Iyengar, M., Schmidt, R. R., Hamann, H., and VanGilder, J., 2007, "Comparison Between Numerical and Experimental Temperature Distributions in a Small Data Center Test Cell," ASME Paper No. IPACK2007-33508.

[4] Iyengar, M., and Schmidt, R. R., 2009, "Analytical Modeling for Thermodynamic Characterization of Data Center Cooling Systems," ASME, J. Electron. Packag., 131(2), p. 021009.

[5] Tsuda, A., Mino, Y., and Nishimura, S., 2017, "Comparison of ICT Equipment Air-Intake Temperatures Between Cold Aisle Containment and Hot Aisle Containment in Datacenters," IEEE International Telecommunications Energy Conference (INTELEC), Broadbeach, QLD, Australia, Oct. 22–26, pp. 59–65.

[6] ASHRAE, 2010, "Save Energy Now Presentation Series," ASHRAE Technical Committee 9.9, Dallas, TX.

[7] Durand-Estebe, B., Le Bot, C., Mancos, J. N., and Arquis, E., 2014, "Simulation of a Temperature Adaptive Control Strategy for an IWSE Economizer in a Data Center," Appl. Energy, 134, pp. 45–56.

[8] Jonge, D. B., 2017, "Trial and Application of Direct Evaporative Cooling at Telstra's Information and Communication Technology Centers," IEEE International Telecommunications Energy Conference (INTELEC), Broadbeach, QLD, Australia, Oct. 22–26, pp. 66–70.

[9] Parolini, L., Sinopoli, B., Krogh, B. H., and Wang, Z., 2012, "A Cyber-Physical Systems Approach to Data Center Modeling and Control for Energy Efficiency," Proc. IEEE, 100(1), pp. 254–268.

[10] Xu, H. G., He, J. P., and Li, Y. Q., 2012, "Energy Management and Control Strategy for DC Micro-Grid in Data Center," IEEE Fifth International Conference on Electricity Distribution, Shanghai, China, Sept. 5–6, pp. 1–6.

[11] Fulpagre, Y., and Bhargav, A., 2015, "Advances in Data Center Thermal Management," Renewable Sustainable Energy Rev., 43, pp. 981–996.

[12] Boucher, T. D., Auslander, D. M., Bash, C. E., Federspiel, C. C., and Patel, C. D., 2006, "Viability of Dynamic Cooling Control in a Data Center Environment," ASME J. Electron. Packag., 128(2), pp. 137–144.

[13] Lin, M., Shao, S., Zhang, X., VanGilder, J. W., Avelar, V., and Hu, X., 2014, "Strategies for Data Center Temperature Control During a Cooling System Outage," Energy Build., 73, pp. 146–152.

[14] Chen, J., Tan, R., Xing, G., and Wang, X., 2014, "PTEC: A System for Predictive Thermal and Energy Control in Data Centers," IEEE Real-Time Systems Symposium, Rome, Italy, Dec. 2–5, pp. 218–227.

[15] Walsh, E. J., Breen, T. J., Punch, J., Shah, A. J., and Bash, C. E., 2011, "From Chip to Cooling Tower Data Center Modeling: Influence of Chip Temperature Control Philosophy," ASME J. Electron. Packag., 133(3), p. 031008.

[16] Shah, A. J., Carey, V. P., Bash, C. E., and Patel, C. D., 2004, "An Exergy-Based Control Strategy for Computer Room Air-Conditioning Units in Data Centers," ASME Paper No. IMECE2004-61384.

[17] Mohsenian, G., Khalili, S., and Sammakia, B., 2019, "A Design Methodology for Controlling Local Airflow Delivery in Data Centers Using Air Dampers," IEEE ITherm Conference, Las Vegas, NV, May 28–31, p. 431.

[18] VanGilder, J., Zhang, Y., Linder, S., and Condor, M., 2019, "Balancing Cooling and IT Airflow With Dampers in Ceiling-Ducted Hot-Aisle Containment in Data Centers," IEEE ITherm Conference, Las Vegas, NV, May 28–31, p. 142.

[19] Baxendale, M., Athavale, J., Robertson, S., and Joshi, Y., 2019, "Data Center Temperature Control Using PI System and MATLAB," IEEE ITherm Conference, Las Vegas, NV, May 28–31, Paper No. 397.

[20] Wemhoff, A. P., del Valle, M., Abbasi, K., and Ortega, A., 2013, "Thermodynamic Modeling of Data Center Cooling Systems," ASME Paper No. IPACK2013-73116.

[21] Steinbrecher, R. A., and Schmidt, R., 2011, "Data Center Environments ASHRAE's Evolving Thermal Guidelines," ASHRAE J., 53(12), pp. 42–49.

[22] Warke, D. A., and Deshmukh, S. J., 2017, "Experimental Analysis of Cellulose Cooling Pads Used in Evaporative Coolers," Int. J. Energy Sci. Eng., 3(4), pp. 37–43.

[23] Joshi, Y., and Kumar, P., 2012, Energy Efficient Thermal Management of Data Centers, 1st ed., Springer, New York, p. 111.

[24] Incropera, F. P., DeWitt, D., Bergman, T., and Lavine, A., 2007, Fundamentals of Heat and Mass Transfer, 6th ed., Wiley, Hoboken, NJ, p. 689.

[25] Bhalerao, A., Fouladi, K., Silva-Llanca, L., and Wemhoff, A. P., 2016, "Rapid Predictions of Exergy Destruction in Data Centers Due to Airflow Mixing," Numer. Heat Transfer, Part A: Appl., 70(1), pp. 48–63.

[26] Grote, K.-H., and Antonsson, E., 2009, "Damper Applications Guide," Handbook of Mechanical Engineering Part B: Applications, Springer, Berlin.

[27] Engineered Software Inc., 2017, "Modeling a Damper," Modeling Piping System Devices, Engineered Software Inc., accessed July 5, 2019, http://kb.eng-software.com/display/ESKB/Modeling+a+Damper

[28] Crane Co. Staff, 2009, Flow of Fluids Through Valves, Fittings and Pipe [Paper No. 410], Crane Company, Stamford, CT, Chap. 6.

[29] Wilcox, S., 2007, "Typical Meteorological Year Weather Database," National Renewable Energy Laboratory, Division of Department of Energy, Golden, CO, Sponsored Under Contract No. DE-AC36-08GO28308, accessed July 5, 2019, http://rredc.nrel.gov/solar/old_data/nsrdb/1991-2005/tmy3/

[30] Brown, K., Torell, W., and Avelar, V., 2014, "Choosing the Optimal Data Center Power Density," Schneider White, Boston, MA, p. 2.

[31] Ibrahim, M., Shrivastava, S., Sammakia, B., and Ghose, K., 2012, "Thermal Mass Characterization of a Server at Different Fan Speeds," 13th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITHERM), San Diego, CA, May 30–June 1, pp. 457–455.

[32] Pardey, Z. M., Demetriou, D. W., Erden, H. S., VanGilder, J. W., Khalifa, H. E., and Schmidt, R. R., 2015, "Proposal for Standard Compact Server Model for Transient Data Center Simulations," ASHRAE Trans., 121(1), pp. 413–421.

[33] Belady, C., Kelkar, K. M., and Patankar, S. V., 1995, "Improved Productivity With Use of Flow Network Modeling (FNM) in Electronic Packaging," Electron. Cooling, 5(1), pp. 36–40.

[34] Wemhoff, A. P., and Frank, A., 2010, "Predictions of Energy Savings in HVAC Systems by Lumped Models," J. Energy Build., 42(10), pp. 1807–1814.

[35] Fried, E., and Idelchik, I. E., 1989, Flow Resistance: A Design Guide for Engineers, 1st ed., Taylor & Francis, New York.

[36] Faul, A. C., 2016, A Concise Introduction to Numerical Analysis, CRC Press, Boca Raton, FL.

[37] White, F., 2001, Fluid Mechanics, 7th ed., McGraw-Hill, New York, p. 769.