Optimized Thermal-Aware Job Scheduling and Control of Data Centers

Tobias Van Damme, Claudio De Persis, and Pietro Tesi

Abstract—Analyzing data centers with thermal-aware optimization techniques is a viable approach to reduce energy consumption of data centers. By taking into account thermal consequences of job placements among the servers of a data center, it is possible to reduce the amount of cooling necessary to keep the servers below a given safe temperature threshold. We set up an optimization problem to analyze and characterize the optimal set points for the workload distribution and the supply temperature of the cooling equipment. Furthermore, under mild assumptions, we design and analyze controllers that regulate the system to the optimal state without knowledge of the current total workload to be handled by the data center. The response of our controller is validated by simulations and convergence to the optimal set points is achieved under varying workload conditions.

Index Terms—Control of constrained systems, cyber-physical systems, Karush-Kuhn-Tucker conditions, Lyapunov methods, optimization and control of data centers.

I. INTRODUCTION

Worldwide energy consumption of data centers reached 350 billion kWh of energy in 2013, 1.73% of the global electricity consumption [2], [3]. With the world being digitized more and more each year, this number is likely to increase as well. Therefore, in the last decade, computer scientists and control engineers have made efforts to reduce the energy consumption of data centers by devising methods to increase the operational efficiency of these computer halls [4].

Although much progress has been made, there are still several challenges in ensuring efficient operation of the cooling equipment. Due to bad design or awareness for the thermal properties of the data center, local thermal hotspots can arise. This causes the cooling equipment to overreact to ensure that the temperature of the equipment stays below the safe thermal threshold. These peaks cause the cooling equipment to consume more energy and then would be necessary if these hotspots were avoided. Therefore, having a good understanding of the thermodynamics involved is vital to increase the cooling efficiency of the data center.

To tackle these challenges, researchers and engineers have studied both software and hardware solutions to this problem. Examples of hardware solutions are isolating cold or hot areas in the data center, or building data centers in cold regions on the planet, where cold outside air can be utilized. Software solutions, on the other hand, focus on strategies that use the knowledge of the thermal properties of the data center to make more intelligent choices how to schedule incoming jobs. Although the two types of solutions are equally important to study, software solutions allow data center operators to implement improvements very fast for very little costs, i.e., implementing new software is less costly than rebuilding a full data center. Furthermore, these types of solutions can provide major performance increases using heuristic methods for smart thermal job schedulers showing up to 30% less energy consumption with respect to nonthermal-aware job schedulers [5], [6]. Therefore, in this paper, we will investigate these thermal-aware software solutions.

Other approaches include considering a heterogeneous data center [7] and using these asymmetric properties to analyze tradeoffs between performance- and energy-aware algorithms, or distinguishing between different types of jobs when scheduling the load [8]. Server consolidation is a natural extension where on top of thermal scheduling, racks are switched ON and OFF to save power. These algorithms usually contain two steps: first to calculate the necessary number of racks and second to calculate the correct workload scheduling [5], [9]–[12].

On the other hand, studies have also been done in a more control theoretical direction. Reference [13] has proposed a control algorithm that tries to maintain the temperature of the equipment around a target value. In [14], it is proposed to implement a two-step algorithm that first minimizes the energy consumption by estimating the required amount of servers to handle the expected workload. In the second step, the algorithm maximizes the response time given a number of servers at its disposal. In an attempt to address scalability, a distributed approach has been studied in [15]. In this paper, units, which range from servers to complete data centers, communicate directly and try to achieve a uniform temperature profile. Another distributed control approach in a hybrid systems setting is proposed in [16]. The hybrid controller tries to evenly divide the total load among the agents in the network in a distributed fashion.

While all these works have strong points on their own, to the best of our knowledge, a thorough formal analysis and
characterization of an energy minimal solution combined with a control strategy, which handles both cooling and job scheduling simultaneously, has not been done before. The objective of this paper is to provide an extendable framework that allows for a characterization of an energy-minimal operating point and then supply methods for operating the data center, such that this operating point is achieved for all load conditions.

The contribution of this paper is twofold. First, from existing thermodynamical principles, we set up a thermodynamical model from which we derive an optimization problem that combines energy minimization with the thermodynamics. In addition to only including temperature constraints [11], we extend the model to also incorporate workload constraints, which allows us to better characterize energy minimal solutions. This design allows for natural extendability to more complicated scheduling policies, such as switching servers ON and OFF.

Second, we develop a novel control strategy for handling the control of the cooling equipment and the workload scheduling simultaneously. Both these control goals have been studied before [13], [17]. However, in [13], the two control goals were handled separately; in [17], a combined algorithm was suggested, but due to complexity could lead to nonoptimal solution. In contrast, our model shows an easy method for handling coordinated cooling and job scheduling control that regulates the system toward the energy optimal solution. Our method is inspired by results from [18], where regulation to optimal steady solutions in the presence of disturbances was considered.

The remainder of this paper is organized as follows. In Section II, the basic thermodynamics are formulated. Then, an optimization problem is formulated in Section III, and under mild assumptions, the equivalence to a reduced form is proven. In Section IV, the optimal solution is analytically analyzed and characterized for different load conditions. Using this analytical solution, a controller is proposed in Section V that can handle unknown load conditions. Finally, in Section VI, a case study is considered to show the performance of the controllers.

Notation: We denote by $\mathbb{R}$ and $\mathbb{R}_{\geq 0}$ the set of real numbers and nonnegative real numbers, respectively. Vectors and matrices are denoted by $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{n \times m}$, respectively. The transpose is denoted by $x^T$, and the inverse of a matrix is denoted by $A^{-1}$. If the entries of $x$ are the functions of time, then the elementwise time derivative is denoted by $\dot{x}(t) := (d/dt)x_i$. By $x_i$, we denote the $i$th element of $x$, and by $a_{ij}$, we denote the $ij$th element of $A$. If a variable already has another subscript, then we switch to superscripts to denote individual elements, i.e., $T^i_{\text{out}}$ and $C^j_\lambda$. We write the diagonal matrix constructed from the elements of vector $x$ as diag$(x_1, x_2, \ldots, x_n)$. The identity matrix of dimension $n$ is denoted by $I_n$, the vector of all ones is denoted by $\mathbb{1} \in \mathbb{R}^n$, and the vector of all zeros is denoted by $\mathbb{0} \in \mathbb{R}^n$. Furthermore, the vector comparison $x \preceq y$ is defined as the elementwise comparison $x_i \leq y_i$ for all elements in $x$ and $y$.

II. System Model

Real life data centers are organized in aisles with many racks each containing a multitude of servers. The cooling of data centers is usually done by air conditioning, and therefore, the racks are set up in a hot- and cold-aisle configuration. Cold air supplied by the computer room air conditioning (CRAC) units is blown into the cold aisles. The air goes through the racks where it absorbs the heat produced by the servers. The air exits the servers in the hot aisle and is recirculated back to the CRAC units, where it is cooled down to the desired supply temperature. A scheduler divides incoming tasks among the racks according to some decision policies. The energy consumption of a rack depends on the amount of tasks it is given. By thermodynamical principles, almost all of this energy consumption is dissipated as heat in the rack. Ideally, all of the exhaust air of the racks is returned to the CRAC; however, due to the complex nature of airflows within the data center, some of the hot air is recirculated back into the cold aisles. This causes the temperature of the air at the inlet of the racks to rise, creating inefficiencies in the cooling of the data center.

It is possible for data centers to have multiple CRAC units. In these cases, we assume that the CRAC units work as one. Allowing different set points will result in mixing of airflows of different temperatures. Airflows of different temperatures, however, are highly nonlinear flows, which depend on the temperature of the flow itself. Therefore, allowing different CRAC set points makes airflows difficult to model and thus adds increased complexity to the already complex situation. This added complexity goes beyond the scope of this paper, and as such, we do not pursue this possibility.

A. Workload

Requests arriving at the data center are collected by a scheduler, which then decides according to some policies how to divide this work among the available racks. We assume that each job has an accompanying tag which denotes the time and the number of computing units (CPU) it requires for execution. Let $J$ denote the integer number of jobs that the scheduler has to schedule in the data center at time $t$. Then, $J(t) = \{1, \ldots, J\}$ denotes the set of jobs to be scheduled at time $t$. Furthermore, let $\lambda_j$ be the number of CPUs that job $j$ requires at time $t$. Then, the total number of CPUs, $D^\star$, that the scheduler has to divide over the racks at time $t$ is given by

$$D^\star(t) = \sum_{j=1}^{J(t)} \lambda_j. \quad (1)$$

We denote by $D_i(t)$ the number of CPUs that the schedulers assign to rack $i$ at time $t$. These variables are collected in the vector

$$D(t) := (D_1(t) \quad D_2(t) \quad \cdots \quad D_n(t))^T.$$  

B. Power Consumption of Racks

Different ways to model power consumption exist, with the main difference being the scope and focus of the models. There are models that try to go as close to the CPU level as possible by modeling the power consumption as a function
of the CPU clock frequency. On the other hand, there are other models that aim at modeling the system on a higher level and capture the power consumption of the CPU as a function of the workload imposed on the server. The models trade between complexity and detail, where the CPU frequency function of the workload imposed on the server. The models level and capture the power consumption of the CPU as other models that aim at modeling the system on a higher level have been found that these models are about 95% accurate [11], [19]–[25]. Therefore, this is our model of choice.

Let \( P_i(t) \) denote the power consumption of rack \( i \) at time \( t \). We model \( P_i(t) \) to consist of a load-independent part, e.g., the equipment consumes a constant amount of power, and a load-dependent part, e.g., the number of CPUs that are actively processing jobs

\[
P_i(t) = \omega_i + u_i D_i(t)
\]

where \( \omega_i \) [W] is the power consumption for the unit being powered on, and \( u_i \) [W CPU\(^{-1}\)] is the power consumption per CPU in use. The variables are collected in the vectors

\[
P(t) := (P_1(t), P_2(t), \ldots, P_n(t))^T
\]

\[
V := (\omega_1, \omega_2, \ldots, \omega_n)^T
\]

and

\[
W := \text{diag}\{\omega_1, \omega_2, \ldots, \omega_n\}
\]

so that

\[
P(t) = V + WD(t).
\]

C. Thermodynamical Model

Following similar arguments as in [13] and [23], a thermodynamical model for each individual rack is constructed. For our model, we focus on the output temperature of the racks as we study the thermodynamical coupling between the workload that is processed by the servers and the energy efficiency of the cooling equipment. As we will show in the following, there is a direct coupling between the output temperature of the racks and both these elements. In Fig. 1, a graphical representation of the heat flows involved is given. The change of temperature of a rack is given by the difference in heat entering and exiting the rack

\[
m_i c_p \frac{d}{dt} T_i^j(t) = Q_i^j(t) - Q_{\text{out}}^j(t) + P_i(t).
\]

Here, \( T_i^j [\text{°C}] \) is the temperature of the exhaust air at rack \( i \), \( c_p [\text{J°C}^{-1} \text{kg}^{-1}] \) is the specific heat capacity of air, \( m_i \) [kg] is the mass of the air inside the rack, \( Q_i^j [\text{W}] \) and \( Q_{\text{out}}^j [\text{W}] \) are the heat entering and exiting the rack, respectively. The heat that enters a rack consists of two parts due to the complex airflows in the data center, i.e., the recirculated air originating from the other racks and the cooled air supplied by the CRAC

\[
Q_i^j(t) = \sum_{j=1}^{n} \gamma_{ji} Q_{\text{out}}^j(t) + Q_{\text{sup}}(t).
\]

Here, \( Q_{\text{sup}} [\text{W}] \) is the heat supplied by the CRAC to rack \( i \) and \( \gamma_{ji} \) is the percentage of the flow, which recirculates from rack \( j \) to rack \( i \). The relation between heat and temperature is given by

\[
Q(t) = \rho c_p f T(t)
\]

where \( \rho \) [kg m\(^{-3}\)] is the density of the air and \( f \) [m\(^3\) s\(^{-1}\)] is the flow rate of the given flow. Combining (5) and (6) with (4) yields

\[
\frac{d}{dt} T_i^j(t) = \frac{\rho}{m_i} \left( \sum_{j=1}^{n} \gamma_{ji} f_j T_i^j(t) - f_i T_i^{j\min}(t) \right) + \frac{\rho}{m_i} \left( f_i - \sum_{j=1}^{n} \gamma_{ji} f_j \right) T_i^{\text{sup}}(t) + \frac{1}{m_i c_p} P_i(t)
\]

where \( T_i^{\text{sup}} [\text{°C}] \) is the temperature of the air supplied by the CRAC and \( f_i \) is the velocity of the airflow through rack \( i \). Rewriting the above relation in the matrix form, i.e., combining the temperature changes of all racks in one equation, results in

\[
\frac{d}{dt} T(t) = A(T(t) - T_{\text{sup}}(t)) + M^{-1} P(t).
\]

Here

\[
T(t) := (T_1^1(t), T_2^1(t), \ldots, T_n^1(t))^T
\]

\[
A := \rho c_p M^{-1} (I^T - I_n) F
\]

\[
F := \text{diag}\{f_1, f_2, \ldots, f_n\}
\]

\[
M := \text{diag}\{c_{p1} m_1, c_{p2} m_2, \ldots, c_{pn} m_n\}
\]

\[
\Gamma := \{\gamma_{ij}\}_{i,j,n}
\]

D. Power Consumption of CRAC

The power consumption of the CRAC is dependent on the temperature of the air, which is returned to CRAC and the supply temperature it has to provide. The airflow, which is returned from rack \( i \) to the CRAC, is given by

\[
f_{\text{ret},i} = \left( 1 - \sum_{j=1}^{n} \gamma_{ij} \right) \frac{\rho}{m_i} f_i
\]
and therefore, the heat returned from all the racks to the CRAC is

\[ Q_{\text{ret}}(t) = \rho c_P \sum_{i=1}^{n} \sum_{j=1}^{n} (1 - \gamma_{ij}) f_i T_{\text{out}}^i(t). \]  

(10)

The heat that the CRAC sends back to the data center is given by \( Q_{\text{sup}}(t) = \rho c_P f_{\text{sup}} T_{\text{sup}}(t) \). With this, the heat that the CRAC has to remove from the air, \( Q_{\text{rem}}(t) \), is given by

\[ Q_{\text{rem}}(t) = Q_{\text{ret}}(t) - Q_{\text{sup}}(t) = \rho c_P \sum_{i=1}^{n} \sum_{j=1}^{n} (1 - \gamma_{ij}) f_i (T_{\text{out}}^i(t) - T_{\text{sup}}(t)) \]

\[ = -T^T \mathbf{MA}(T_{\text{out}}(t) - T_{\text{sup}}(t)). \]  

(11)

To determine the amount of work that the CRAC has to do to remove a certain amount of heat, Moore et al. [5] introduced the coefficient of performance, COP\((T_{\text{sup}}(t))\), to indicate the efficiency of the CRAC as a function of the target supply temperature. They found that CRAC units work more efficiently when the target supply temperature is higher. The COP represents the ratio of heat removed to the amount of work necessary to remove that heat. For a water-chilled CRAC unit in the HP Utility Data Center, they found that the COP is a quadratic, increasing function. In a general sense, the COP can be any monotonically increasing function. The power consumption of the CRAC units can then be given by

\[ P_{\text{AC}}(T_{\text{out}}(t), T_{\text{sup}}(t)) = \frac{Q_{\text{rem}}(t)}{\text{COP}(T_{\text{sup}}(t))}. \]  

(12)

Assumption 1: The function \( \text{COP}(T_{\text{sup}}) \) of the CRAC unit considered in this paper is monotonically increasing in the range of operation for \( T_{\text{sup}} \).

Example 1: Let us consider a small example to illustrate the influence of a small difference in supply temperature on the power consumption of the CRAC. Consider the quadratic COP\((T_{\text{sup}}(t))\) found in [5], and two cases where the returned air has to be cooled down by 5 °C, in the first case from 25 °C to 20 °C and in the second case from 30°C to 25°C. Assume that the energy contained in 5 °C temperature difference of air is 100 W. In the first case, COP(20) = 3.19, and in the second case, COP(25) = 4.73. By (12), the energy consumed by the CRAC to cool down the returned air to the required temperature is

\[ P_{\text{AC1}} = \frac{100}{3.19} = 31.34 \text{ W}, \quad P_{\text{AC2}} = \frac{100}{4.73} = 21.14 \text{ W}. \]

Here, it is shown that if the temperature of the returned air increases by 5 °C, the power consumption of the CRAC unit decreases by 30%.

Having completed the model finally allows us to formulate the control problem we would like to solve.

### III. Problem Formulation

The objective of this paper is twofold: First, we want to find optimal set points for the temperature distribution, supply temperature, and workload distribution that minimize the power consumption of the data center. Second, we want to design controllers that ensure convergence of the variables to the obtained set points. Hence, the control problem is defined as follows.

**Problem 1:** For system (8), design controllers for the workload distribution \( D(t) \) and supply temperature \( T_{\text{sup}}(t) \), such that, given an unmeasured total load \( D^* \), any solution of the closed-loop system is bounded and satisfies

\[ \lim_{t \rightarrow \infty} (T_{\text{out}}(t) - \bar{T}_{\text{out}}) = 0 \]  

(13)

\[ \lim_{t \rightarrow \infty} (T_{\text{sup}}(t) - \bar{T}_{\text{sup}}) = 0 \]  

(14)

\[ \lim_{t \rightarrow \infty} (D(t) - \bar{D}) = 0 \]  

(15)

where \( \bar{T}_{\text{out}}, \bar{T}_{\text{sup}}, \) and \( \bar{D} \) are the optimal set-point values for the temperature distribution, supply temperature, and power consumption, i.e., workload distribution, respectively, which are defined in Section III-A.

From this point on, we will implicitly assume the dependence of the variables on time and only denote it when confusion might arise otherwise.

#### A. General Optimization Problem

We formulate an optimization problem to minimize the power consumption while taking into account the physical constraints of the equipment, i.e., the servers only have finite computational capacity and the temperature of the servers cannot exceed a certain threshold. The power consumption of the data center can be written as a combination of two parts, the power consumption of the cooling equipment and the power consumption of the racks. Combining (3) and (12), we can write the total power consumption as

\[ C(T_{\text{out}}, T_{\text{sup}}, D) = \frac{Q_{\text{rem}}}{\text{COP}(T_{\text{sup}})} + T^T P(D). \]  

(16)

A reasonable way [11], [14] to formulate the optimization problem is

\[ \min_{T_{\text{out}}, T_{\text{sup}}, D} \frac{Q_{\text{rem}}}{\text{COP}(T_{\text{sup}})} + T^T P(D) \]  

(17a)

\[ \text{s.t. } D^* = T^T D \]  

(17b)

\[ 0 \leq D \leq D_{\text{max}} \]  

(17c)

\[ 0 = A(T_{\text{out}} - T_{\text{sup}}) + M^{-1} P(D) \]  

(17d)

\[ T_{\text{out}} \leq T_{\text{safe}}. \]  

(17e)

Equation (17b) ensures that all the available works are divided among the racks, and (17c) encompasses the computational capacity of the rack, i.e., rack \( i \) has \( D_{\text{max}} \) CPUs available at most. The system dynamics should be at steady state once the optimal point has been reached [see (17d)], and finally, (17e) enforces that the temperature of the racks is below the given safe threshold, \( T_{\text{safe}} \in \mathbb{R}^n \).

#### B. Equivalent Optimization Problem for Homogeneous Data Centers

Due to the nonlinear nature of how the COP affects the power consumption, it is not trivial to analyze the general optimization problem. Although (17) is a difficult problem...
to solve analytically, it is possible to reduce the optimization problem to a simpler equivalent problem for a specific important case. In many of the larger real-life data centers, most of the equipment is identical, i.e., the power consumption characteristics of the computational equipment is identical, that is \( v_i = v \) and \( w_j = w \) for all \( i \) in (2). It is desirable for data centers to employ identical equipment, because this allows for decreased maintenance complexity and allows for bulk purchases of the equipment, which reduce operational costs. In this case, the data center is said to be composed of homogeneous racks or, more simply, the data center is homogeneous.

In case of a homogeneous data center, the power consumption is given by \( P(D) = v \mathbb{1} + w D \), and the total computational power consumption is given by

\[
\mathbb{1}^T P(D) = \nu \mathbb{1} + w \mathbb{1}^T D = \nu \mathbb{1} + w D. \quad (18)
\]

For this case, the computational power consumption no longer depends on the way that the jobs are distributed but only depends on the total workload. This property simplifies the cost function defined in (16) considerably.

**Theorem 1:** Let the data center consist of homogeneous racks, i.e., \( v_i = v \) and \( w_j = w \) for all \( i \) in (2) and assume that constraint (17d) is satisfied. Then, problem (17) is equivalent to

\[
\begin{align*}
\max_{T_{\text{out}}} & \quad C_1^T T_{\text{out}} \\
\text{s.t.} & \quad 0 \preceq C_3 T_{\text{out}} + C_4(D^*) \preceq D_{\text{max}} \\
& \quad T_{\text{out}} \preceq T_{\text{safe}},
\end{align*}
\]

for suitable \( C_1, C_3 \) and \( C_4 \).

Before we prove this theorem, we need to introduce some notation and extra preparatory results. In these preparatory results (Lemmas 1–3), the homogeneity condition is not required, and statements are given in terms of the power consumption vector \( P \) defined as in (3).

**Lemma 1:** Equation (17d) implies that the following relation holds:

\[
\mathbb{1}^T P(D) = -\mathbb{1}^T M (T_{\text{out}} - T_{\text{sup}}) = Q_{\text{rem}}
\]

with \( Q_{\text{rem}} \) defined in (11), which reduces the cost function to

\[
C(T_{\text{out}}, T_{\text{sup}}, D) = \left( 1 + \frac{1}{\text{COP}(T_{\text{sup}})} \right) \mathbb{1}^T P(D). \quad (20)
\]

**Proof:** By multiplying (17d) by \( \mathbb{1}^T M \) and solving for \( \mathbb{1}^T P(D) \), we obtain the above result.

**Lemma 2:** If (17b) and (17d) are satisfied, then

\[
\begin{align*}
T_{\text{sup}} &= C_1^T T_{\text{out}} + C_2(D^*) \\
C_1^T &= \frac{\mathbb{1}^T W^{-1} MA}{\mathbb{1}^T W^{-1} M} \\
C_2(D^*) &= \frac{D^* + \mathbb{1}^T W^{-1} V}{\mathbb{1}^T W^{-1} M}.
\end{align*}
\]

**Proof:** After multiplying (17d) by \( \mathbb{1}^T W^{-1} M \), combining with (17b) and some basic matrix manipulations, the result is obtained.

**Lemma 3:** If (17b) and (17d) are satisfied, then

\[
\begin{align*}
D &= C_3 T_{\text{out}} + C_4(D^*) \\
C_3 &= \frac{W^{-1} MA (I_n - \mathbb{1} C_1^T)}{\mathbb{1}^T W^{-1} M} \\
C_4(D^*) &= \frac{W^{-1} MA \mathbb{1} C_2(D^*) - W^{-1} V}{\mathbb{1}^T W^{-1} M}.
\end{align*}
\]

**Proof:** Substituting the result of Lemma 2 in (17d), premultiplying (17d) by \( W^{-1} M \), and solving for \( D \) yield the result.

**Remark 1:** The dimensions of the constants from Lemmas 1–3 are \( C_1 \in \mathbb{R}^n, C_2 \in \mathbb{R}, C_3 \in \mathbb{R}^{n \times n} \), and \( C_4 \in \mathbb{R}^n \). The following identities for the constants \( C_1, C_3, \) and \( C_4 \) are observed:

\[
C_1^T \mathbb{1} = 1, \quad C_3 \mathbb{1} = 0, \quad C_4 \mathbb{1} = D^* \quad (23)
\]

An important consequence worth to note is that the constant \( \mathbb{1}^T D \), with \( D \) defined as in (22), satisfies the identity \( \mathbb{1}^T D = D^* \).

Lemmas 2 and 3 show that at the steady state, the supply temperature \( T_{\text{sup}} \) and workload distribution vector \( D \) are uniquely defined by the total workload \( D^* \) and the temperature distribution \( T_{\text{out}} \). With these properties in mind, we are ready to prove Theorem 1.

**Proof of Theorem 1:** Assume that Problem (17) has a solution. By Lemma 1, the cost function reduces to (20). By the homogeneity assumption, (18) holds, which shows that the cost function (20) is independent of the distribution \( D \) and depends only on \( T_{\text{sup}} \). Hence, in view of Assumption 1 (monotonicity of the function \( \text{COP}(T_{\text{sup}}) \)), a solution to Problem (17) is the one that maximizes \( T_{\text{sup}} \). By (21) in Lemma 2, this solution must maximize the cost function in (19a). The constraints in (17) and Lemma 3 imply the constraints in (19), showing that a solution to (17) must also be a solution to (19). Conversely, if a solution to (19) exists, define \( D \) as in (22), and notice that (17b) is satisfied, as it is promptly verified using the identities (23). Then, by the homogeneity assumption, (17d), Lemma 1, and Lemma 2, maximizing the cost function in (19a) implies minimizing the cost function in (17a). Moreover, the definition of \( D \) and the constraint (19b) implies (17c). Constraint (17e) trivially holds because of (19c).

This ends the proof.

**IV. CHARACTERIZATION OF THE OPTIMAL SOLUTION**

In Section III, we have shown the possibility to reduce the optimization problem to a simpler form. In this section, we show that using Karush-Kuhn-Tucker (KKT) optimality conditions, it is possible to further characterize the optimal point.

**A. KKT Optimality Conditions**

Because the optimization problem (19) is convex and all inequality constraints are linear functions, we have that Slater’s condition holds. Therefore, it follows that \( T_{\text{out}} \) is an optimal solution to (19) if and only if there exists \( \bar{\mu}, \bar{\mu}^+, \bar{\mu}^- \in \mathbb{R}_{\geq 0}^n \).
such that the following set of relations is satisfied \[26\]:

\[-C_1 + \bar{\mu} + C_1^T (\bar{\mu}_+ - \bar{\mu}_-) = 0,\]  
\[0 \preceq C_3 \bar{T}_{out} + C_4 (D^*) \preceq D_{\text{max}},\]  
\[\bar{T}_{out} \preceq T_{\text{safe}},\]  
\[\bar{\mu}_+^T (C_3 \bar{T}_{out} + C_4 (D^*)) = 0,\]  
\[\bar{\mu}_-^T (-C_3 \bar{T}_{out} - C_4 (D^*)) = 0,\]  
\[\bar{\mu}_+^T (\bar{T}_{out} - T_{\text{safe}}) = 0,\]  
\[\bar{\mu}_+^T + \bar{\mu} - \bar{\mu}_- \succ 0.\] (24g)

The Lagrangian corresponding to the optimal problem is given by

\[\mathcal{L}(\mu, \mu_+, \mu_-, \bar{T}_{out}) = -C_1^T T_{out} + \mu^T (\bar{T}_{out} - T_{\text{safe}}) + \mu_+^T (-C_3 T_{out} - C_4 (D^*)) + \mu_-^T (C_3 T_{out} + C_4 (D^*)) - D_{\text{max}}.\] (25)

**B. Characterization of Optimal Temperature Profile**

By studying the KKT optimality conditions, we can characterize the optimal solution in different cases.

1) **Inactive Workload Constraints:** Every rack is processing some work, but not all the processors of each rack are utilized

\[0 < (C_3 \bar{T}_{out} + C_4 (D^*))_i < D_{\text{max}}^i \forall i \in \{1, \ldots, n\}.\]

2) **Partially Active Workload Constraints:** In \(k\) racks, all processors are processing jobs. The other \(n-k\) racks are processing some work, but still have processors available

\[(C_3 \bar{T}_{out} + C_4 (D^*))_i = D_{\text{max}}^i \forall i \in \{1, \ldots, k\},\]

\[0 < (C_3 \bar{T}_{out} + C_4 (D^*))_i < D_{\text{max}}^i \forall i \in \{k+1, \ldots, n\}.\]

The optimal temperature profile corresponding to these two cases is summarized in Theorems 2 and 3.

**Theorem 2:** Assume the case that none of the workload constraints are active, i.e.,

\[0 < (C_3 \bar{T}_{out} + C_4 (D^*))_i < D_{\text{max}}^i \forall i \in \{1, \ldots, n\}.\]

The solution to (24) and the optimal solution for the optimization problem (19) is then given by

\[\bar{\mu}_+ = \bar{\mu}_- = 0, \quad \bar{\mu} = C_1 > 0, \quad \bar{T}_{out} = T_{\text{safe}}.\] (26)

**Proof:** Because all the inequality constraints regarding the workload are inactive, we have that \(C_3 \bar{T}_{out} + C_4 (D^*) - D_{\text{max}} < 0\) and \(-C_3 \bar{T}_{out} - C_4 (D^*) < 0\). Then, from (24d) and (24e), we have that \(\bar{\mu}_+ = \bar{\mu}_- = 0\). From (24a), it follows that \(\bar{\mu} = C_1 > 0\), such that from (24f), we conclude that \(\bar{T}_{out} = T_{\text{safe}}.\)

**Theorem 3:** In the case that a part of the workload constraints are active, i.e.,

\[(C_3 \bar{T}_{out} + C_4 (D^*))_i = D_{\text{max}}^i \forall i \in \{1, \ldots, k\},\]

\[0 < (C_3 \bar{T}_{out} + C_4 (D^*))_i < D_{\text{max}}^i \forall i \in \{k+1, \ldots, n\},\]

the solution of (24) is as follows.

1) For the racks at the constraint boundary, \(i \in \{1, \ldots, k\}\)

\[\bar{\mu}_- = 0, \quad C_1^i + \sum_{j=1, j \neq i}^{k} \bar{\mu}_+^j |C_{3j}^i| \geq \bar{\mu}_+^i \geq 0\] (27)

\[\bar{\mu}_+^i = C_1^i + \sum_{j=1, j \neq i}^{k} \bar{\mu}_+^j |C_{3j}^i| - \bar{\mu}_+^i C_{3j}^i \geq 0 \quad \forall i \in \{k+1, \ldots, n\}.\] (28)

\[\bar{T}_{out}^i = D_{\text{max}}^i C_3^i + \sum_{j=k+1}^{n} |C_{3j}^i| T_{\text{safe}}^j + \sum_{j=1, j \neq i}^{k} \bar{\mu}_+^j C_{3j}^i \leq T_{\text{safe}}.\] (29)

2) For the racks that are not at the constraint boundary, \(i \in \{k+1, \ldots, n\}\)

\[\bar{\mu}_-^i = \bar{\mu}_+^i = 0\]

\[\bar{\mu}_+^i = C_1^i + \sum_{j=1}^{k} \bar{\mu}_+^j |C_{3j}^i| > 0 \quad \forall i \in \{k+1, \ldots, n\}.\]

\[\bar{T}_{out}^i = T_{\text{safe}}^i.\] (32)

Before we can prove Theorem 3, we need to know more about the structure of \(C_3\).

**Property 1:** Consider \(C_3\) as defined in Lemma 3. The off-diagonal terms of this matrix are strictly negative and the diagonal terms are strictly positive.

**Proof:** The proof can be found in Appendix A.

**Proof of Theorem 3:** Because a part of the workload constraints are at the constraint boundary, the analysis following from the Lagrange multipliers is more involved. First, we can say that

\[\bar{\mu}_- = 0 \quad \forall i \in \{1, \ldots, k\},\]

\[\bar{\mu}_+ = 0 \quad \forall i \in \{k+1, \ldots, n\},\]

\[\bar{\mu}_+^i \geq 0 \quad \forall i \in \{1, \ldots, k\}.\]

Then, from (24a)

\[\bar{\mu}_+ = C_1^i - \sum_{j=1}^{k} \bar{\mu}_+^j C_{3j}^i.\] (33)

From Property 1, we have that the off-diagonal elements of \(C_3\) are strictly negative. For racks \(i \in \{k+1, \ldots, n\}\), we have that the \(C_{3j}^i\) elements in (33) will always be off-diagonal elements. Therefore, rewriting (33) gives

\[\bar{\mu}_+^i = C_1^i + \sum_{j=1}^{k} \bar{\mu}_+^j |C_{3j}^i| > 0 \quad \forall i \in \{k+1, \ldots, n\}.\] (34)

and then from (24f), it holds that

\[\bar{T}_{out}^i = T_{\text{safe}}^i \quad \forall i \in \{k+1, \ldots, n\}.\] (35)

For racks \(i \in \{1, \ldots, k\}\), (33) is given by

\[\bar{\mu}_+^i = C_1^i + \sum_{j=1, j \neq i}^{k} \bar{\mu}_+^j |C_{3j}^i| - \bar{\mu}_+^i C_{3j}^i \geq 0.\] (36)
For (36) to hold, it should hold that
\[ \frac{C_i^j + \sum_{j=1,j \neq i}^k \bar{R}_i^j |C_i^j|}{C_i^g} \geq \bar{\mu}_i^j \quad \forall i \in \{1, \ldots, k\}. \] (37)
As the left-hand side of (37) is strictly positive for all \( i \in \{1, \ldots, k\} \), it is possible to find feasible \( \bar{\mu}_i^j \geq 0 \), such that \( \bar{\mu}_i^j \geq 0 \) for all \( i \). It can be shown that \( \bar{T}_{\text{out}} \) for all \( i \in \{1, \ldots, k\} \) is given as
\[
\bar{T}_{\text{out}}^i = \frac{D_{\text{max}} - C_i^j(D^*)}{C_i^g} + \sum_{j=k+1}^n \frac{|C_i^j|}{C_i^g} T_{\text{safe}}^j \\
+ \sum_{j=1,j \neq i}^k \frac{|C_i^j|}{C_i^g} T_{\text{out}}^j \leq T_{\text{safe}}^i. \] (38)

Remark 2: One cannot freely choose the \( k \) racks for which \( D_i = D_{\text{max}} \). Whether or not a rack is processing its maximum capacity depends on the data center parameters, i.e., small amount of recirculated air at the input of the rack and low power consumption of the computational equipment. For these racks, it holds that
\[
\bar{T}_{\text{out}}^i \leq T_{\text{safe}}^i \quad \forall i \in \{1, \ldots, k\}. \]

V. TEMPERATURE-BASED JOB SCHEDULING CONTROL

As established in Section IV, it is possible to calculate the optimal solution under the assumption that the total workload at time \( t \), \( D^* \) is known. However, it might not always be possible to obtain this quantity. For example when jobs arrive in the data center in some cases, it might be hard to assess how much resources these jobs need. Consider the case where a virtual machine is requested by a user. Usually, a certain amount of resources are allocated to this virtual machine; however, the users need not use all the available resources all the time. In those situations, it is hard to obtain the real workload. In this section, we design a controller that is still valid in those setups. Since \( D^* \) is constant and \( \bar{T}_{\text{out}}(0) = D^* \). Then, the solution of system (8) with controllers (39) and (40) is bounded and converges to the optimal solution of the problem defined in (17) and therefore satisfies control objectives (13)–(15).

Proof: For ease of notation, we introduce incremental variables to denote deviations from optimal values
\[
\tilde{T}_{\text{out}} = T_{\text{out}} - \bar{T}_{\text{out}} \\
\tilde{T}_{\text{sup}} = T_{\text{sup}} - \bar{T}_{\text{sup}} \\
\tilde{D} = D - \bar{D}
\]
where \( \tilde{T}_{\text{out}} = T_{\text{safe}}, \tilde{T}_{\text{sup}} \) as in (21) and \( \tilde{D} \) defined as the right-hand side of (22). With these definitions, system (8) can be rewritten as
\[
\dot{\tilde{T}}_{\text{out}} = A \tilde{T}_{\text{out}} - A \tilde{T}_{\text{sup}} + B \tilde{D} \] (42)
where \( A \) and \( B \) are as before
\[
A = \rho \bar{C}_p M^{-1}(\Gamma^T - I_n)F \\
B = M^{-1}W.
\]
Define the incremental storage functions as
\[
\Xi_1(\tilde{T}_{\text{sup}}) = \frac{1}{2} \| \tilde{T}_{\text{sup}} \|^2 \] (43)
\[
\Xi_2(\tilde{D}) = \frac{1}{2} \| \tilde{D} \|^2. \] (44)
The storage functions satisfy
\[
\dot{\Xi}_1(\tilde{T}_{\text{sup}}, \tilde{T}_{\text{out}}) = \tilde{T}_{\text{sup}}^T \dot{\tilde{T}}_{\text{sup}} = \tilde{T}_{\text{sup}}^T \dot{A}^T Z T_{\text{out}} \] (45)
and
\[
\dot{\Xi}_2(\tilde{D}, \tilde{T}_{\text{out}}) = \dot{\tilde{D}}^T \tilde{D} = \tilde{D}^T \left( \frac{T_{\text{out}}^T}{n} - I_n \right) B^T Z T_{\text{out}} \] (46)
\[
= \tilde{D}^T \frac{T_{\text{out}}^T}{n} B^T Z T_{\text{out}} - \tilde{D}^T B^T Z T_{\text{out}}. \] (47)
Note that $1^T D(t) = D^*$ is satisfied for all $t \geq 0$. In fact, first we notice that $1^T D = 0$ at all times $t \geq 0$. Hence, if $1^T D(0) = D^*$, then $1^T D(t) = D^*$ for all $t \geq 0$. With this, we see that $\tilde{D}^T \| = (D - \bar{D})^T \| = D^* - D^* = 0$, such that (47) is reduced to

$$\hat{z}_2(\tilde{D}, \tilde{T}_{out}) = - \tilde{D}^T B^T Z \tilde{T}_{out}. \quad (48)$$

Now, consider the following Lyapunov function $V(\tilde{T}_{out}) = \frac{1}{2} \tilde{T}_{out}^T Z \tilde{T}_{out}$, where $Z$ is defined in (41). Then, $V(\tilde{T}_{out})$ satisfies

$$\dot{V}(\tilde{T}_{out}) = - \| \tilde{T}_{out} \|^2 - \tilde{T}_{sup}^T A^T Z \tilde{T}_{out} + \tilde{D}^T B^T Z \tilde{T}_{out}. \quad (49)$$

Hence, the total Lyapunov function $\Xi_1 + \Xi_2 + V$ satisfies

$$\dot{V}(\tilde{T}_{out}) + \hat{z}_1(\tilde{T}_{sup}, \tilde{T}_{out}) + \hat{z}_2(\tilde{D}, \tilde{T}_{out}) = - \| \tilde{T}_{out} \|^2 \leq 0. \quad (50)$$

Since $\Xi_1 + \Xi_2 + V$ is radially unbounded, (50) implies the boundedness of the solutions. Using LaSalle’s invariance principle, this result implies that every solution to the closed-loop system initialized as $1^T D(0) = D^*$ converges to the largest invariant set, where $\tilde{T}_{out} = 0$. Next, we show that $\tilde{D}$ and $\tilde{T}_{sup}$ are zero on this invariant set. Because $\tilde{T}_{out}$ is zero, (42) reduces to

$$0 = -A \tilde{T}_{sup} + B \tilde{D}. \quad (51)$$

Premultiplying this by $1^T B^{-1}$, we get

$$-1^T \tilde{D} = 0 = -1^T B^{-1} A \tilde{T}_{sup} \quad (52)$$

and since

$$-1^T B^{-1} A \| > 0 \quad (53)$$

we obtain that $\tilde{T}_{sup} = 0$. To understand why (53) holds true, observe that $A \|$ has all entries strictly negative, as it is immediately deduced from (63) and (64) in Appendix B. Now, the inequality easily follows.

With $\tilde{T}_{sup} = 0$ and with $B$ nonsingular, it follows from (51) that $\tilde{D} = 0$. Hence, the largest invariant set to which the solutions converge is the singleton $\{ \tilde{T}_{out}, \tilde{T}_{sup}, \tilde{D}, \} = (0, 0, 0, 0)$. Therefore, we conclude that system (42) with controllers (39) and (40) satisfies control objectives (13)–(15), and the state and the inputs of the system converge to the optimal solution.

The proposed controller for the workload rebalances the workload currently present in the data center. The initial scheduling is assumed to be taken care of by an external entity over which we have no control. This approach is most applicable in cases where the initial scheduling is done in a noncontrollable way, e.g., when the scheduling is hard-coded and incoming jobs are scheduled by means of chassis numbers. In these situations, the only option available is to move jobs around to drive the data center to the optimal state.

The above-mentioned result shows the guaranteed asymptotic tracking of constant reference signals. However, in practice, the controller can handle variations in set points, provided that the set points change sufficiently slow. In Section VI, we will study the behavior of our controller under varying set points in a real data center context.

VI. Case Study

To evaluate the performance of the proposed controller, we use MATLAB to simulate the closed-loop system with a synthetic workload trace. For both the data center parameters and the workload trace, we use the data presented in [13]. The data center parameters were obtained from measurements by Vasic et al. [13] at the IBM Zurich Research Laboratory. These data are to the best of knowledge the most extensive characterization of the heat recirculation parameters of a data center.

A. Data Center Parameters

The simulated data center consists of 30 homogeneous server racks, i.e., the power consumption characteristics, safe temperature threshold, and physical parameters are identical for all 30 racks. The rack model is a Dell PowerEdge 1855, with 10 dual-processor blade servers, i.e., a total of 20 CPU units per rack. The power consumption of the racks is modeled by $P_i(t) = 1728 + 145.5 D_i(t)$ [12]. The safe threshold temperature is set at $30^\circ$C. We supply a synthetic workload trace to the data center (see Fig 2). The workload trace is constructed by varying the total workload by $\pm 10\%$ about two nominal values, representing nighttime and daytime operation levels. The total workload changes every 7.5 min during which the workload is assumed to be constant.

![Fig. 2. Synthetic workload trace supplied to data center. The workload varies $\pm 10\%$ about two nominal values, representing nighttime and daytime operation levels. The total workload changes every 7.5 min during which the workload is assumed to be constant.](image)
The response of $(T_{\text{out}} - \bar{T}_{\text{out}})$ during the simulation for four selected racks. The full simulation is shown in the inset, and the main plot is a magnification of the response after a change in total workload around $t = 10$ hours. Each time the total workload changes, the temperature of the racks starts to deviate from the optimal value and the controllers drive the data center to the new optimal solution, $(T_{\text{out}} - \bar{T}_{\text{out}}) = 0$ again. The oscillatory response of the output temperature coincides with the response of the supply temperature controller. Over the whole simulation, the temperature is kept in a bandwidth of $\pm 0.5 \, ^\circ\text{C}$ around the target temperature distribution.

The response of $(T_{\text{sup}} - \bar{T}_{\text{sup}})$ during the simulation for four selected racks. The full simulation is shown in the inset, and the main plot is a magnification of the response after a change in total workload around $t = 10$ hours. The initial overshoot depends on the change of the total workload, i.e., the difference between the optimal supply temperatures in the two intervals. The oscillatory response results in a fluctuating output temperature profile.

The response of $(T_{\text{out}} - \bar{T}_{\text{out}})$ during the simulation for four selected racks. The full simulation is shown in the inset, and the main plot shows the temperature response over a larger time interval, which covers multiple changes in total workload. The fast response of the controllers is clearly visible here, and we see that, after a very short transient, the controllers steer the temperature of the servers back to the optimal value.

The supplied workload simulated a day and night cycle to study the response of the controller under large varying loads. From the results, we see no difficulty for the controller to handle these different conditions. We conclude that the controller is able to keep the temperature of the racks around the target set point under all load conditions.

VII. CONCLUSION AND FUTURE WORK

Many papers on thermal-aware job scheduling have studied the topic from a practical perspective; however, a theoretical analysis has less often been done. In this paper, we describe data centers and corresponding thermodynamics in a control theoretical fashion combining the optimization theory with a controller design.
We have studied the minimization of energy consumption in a data center, where recirculation of airflow is present, i.e., inefficiencies in cooling of the racks, through thermal-aware job scheduling and cooling control. We have set up an optimization problem and characterized the optimal workload distribution and cooling temperature to achieve minimal energy consumption while ensuring job processing and thermal threshold satisfaction. In addition, we have presented controllers that track a reference signal and are able to drive the control and state variables to the optimal values. Furthermore, simulations show that the controllers can work with varying workload conditions, as the convergence time of the controllers is significantly faster than the frequency of the workload variation.

We have shown that it is possible to uniquely determine the optimal cooling supply temperature and workload distribution as a function of the total workload and desired temperature distribution of the racks in the data center. Furthermore, we have shown that the optimal temperature distribution can be analytically calculated and that this distribution is independent of the workload distribution if none of the racks reach its computational capacity.

With the assumption that none of the racks are at its computational capacity, we have designed controllers that control the supply temperature and workload distribution to drive the data center to the optimal state.

There are several directions in which we want to extend our research. First, we want to extend the framework to include situations where the optimal temperature distribution changes due to racks reaching their computational capacity. This will allow us to include server consolidation, where the number of active racks is decreased to further reduce energy consumption. In these situations, it is inevitable that the computational capacity of some of the racks is reached and that varying optimal temperature distributions will have to be addressed.

Our control approach requires knowledge of the thermal characteristics of the data center. Studying the robustness and stability of our approach under small variations of the heat recirculation matrix is, therefore, of importance. Last, it would be interesting to study the possibility of allowing multiple CRAC units with different set points, or including other variables in the optimization problem, such as service-level agreements and response times of the jobs.

**APPENDIX A**

**Proof of Property 1**

From Lemma 3, we have that

\[ C_3 = -W^{-1}MA(I_n - C_1^T) \]

where

\[ C_1^T = \frac{1}{\omega} W^{-1}MA \frac{1}{\omega} \]

Defining a temporary variable \( \alpha = W^{-1}MA \), we can write \( C_3 \) as

\[ C_3 = -\alpha + \frac{1}{\omega^2} \alpha l^T \alpha. \]

The \( ij \)-th component of \( C_3 \) is then given by

\[ C_{3ij} = -\alpha_{ij} + \frac{1}{\omega} \sum_{l=1}^{n} \alpha_{il} \sum_{k=1}^{n} \alpha_{kj} \frac{1}{\omega}. \]  

From the definition of \( \alpha \), we find that the \( ij \)-th component of \( \alpha \) is given by

\[ \alpha_{ij} = c_p \rho \frac{1}{\omega w} (\gamma_{ji} - \delta_{ji}) f_j \]

where \( \delta_{ji} \) is the Kronecker delta, which is 1 if \( i = j \) and 0 otherwise. To simplify the mathematics a little from now on, we assume that the data center consists of homogeneous racks [see (18)]. Combining (55) with (54), we have

\[ C_{3ij} = -c_p \rho \frac{1}{\omega w} \left( (\gamma_{ji} - \delta_{ji}) f_j + \sum_{l=1}^{n} \gamma_{il} f_l \sum_{k=1}^{n} \gamma_{jk} f_k \right). \]

Although the big fraction in (56) looks a bit daunting, it is actually easy to conceptually understand it. The airflow at the inlet of the rack consists of two parts, air coming from the CRAC unit and air recirculating from other racks to the rack in question. At the outlet of the rack, the airflow is composed of the air going back to the CRAC unit and the air recirculating from the rack in question to all the other racks. Looking closer at the numerator of (56), we see that the first half is the air flowing from the CRAC unit to rack \( i \), and the second half is the air flowing from rack \( j \) to the CRAC unit. The denominator is the sum of the airflow that each rack receives from the CRAC unit, which is equal to the supplied airflow, \( f_{\text{sup}} \). In this way, we can simplify (56) to

\[ C_{3ij} = -c_p \rho \frac{1}{\omega w} \left( (\gamma_{ji} - \delta_{ji}) f_j + \sum_{l=1}^{n} \gamma_{il} f_l \right) \frac{f_{\text{CRAC to } i} f_{\text{CRAC to } i}}{f_{\text{sup}}} \]

Now, in the case that \( i \neq j \), (57) is reduced to

\[ C_{3ij} < 0. \]

Here, we see that the off-diagonal terms of \( C_3 \) are strictly negative.

As for the diagonal terms, \( i = j \), we have

\[ C_{3ii} = c_p \rho \frac{1}{\omega w} \left( 1 - \gamma_{ii} f_i + \sum_{l=1}^{n} \gamma_{il} f_l \right) \frac{f_{\text{CRAC to } i}}{f_{\text{sup}}} \]

Since

\[ (1 - \gamma_{ii}) f_i = f_i - \sum_{l=1}^{n} \gamma_{il} f_l \sum_{l=1, l \neq i}^{n} \gamma_{il} f_l \]

(59)
we have that

\[
C_{3ij} = c_{p} \rho \frac{1}{\mu} \left( \sum_{l=1, l \neq i}^{n} \gamma_{li} f_{l} + f(\text{CRAC to } i) \left( 1 - \frac{f_{i} \text{(to CRAC)}}{f_{\sup}} \right) \right)_{>0}
\]

> 0. \tag{61}

In (61), we see that the diagonal terms of \(C_3\) are strictly positive. This concludes the proof. \(\square\)

**APPENDIX B**

**PROOF OF HURWITZ PROPERTY OF MATRIX A**

Matrix \(A\) as defined in Section II-C is given by

\[
A = \rho c_{p} M^{-1} (I^{T} - I_{h}) F.
\]

Writing the matrix out in full gives

\[
A = \rho \left( \begin{array}{cccc}
\gamma_{11} - 1 & \frac{\gamma_{21}}{m_1} & \frac{\gamma_{22}}{m_2} & \cdots & \frac{\gamma_{n1}}{m_n} \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\frac{\gamma_{1n}}{m_n} & \frac{\gamma_{2n}}{m_n} & \cdots & \cdots & \frac{\gamma_{nn} - 1}{m_n} \\
\end{array} \right) \tag{62}
\]

If we can show that matrix \(A\) is strictly diagonal dominant and that the diagonal elements are negative, then, by the Gershgorin circle theorem, we have shown that matrix \(A\) is Hurwitz.

First, we will prove the strict diagonal dominance of matrix \(A\). As stated in Appendix A, the airflow in a rack consists of two parts, the recirculated air from the other racks and the supplied air by the CRAC, namely,

\[
f_{i} = \gamma_{ii} f_{i} + \sum_{j=1, j \neq i}^{n} \gamma_{ij} f_{j} + f_{\sup}.
\]

Hence

\[
(\gamma_{ii} - 1) f_{i} = - \sum_{j=1, j \neq i}^{n} \gamma_{ij} f_{j} - f_{\sup} < - \sum_{j=1, j \neq i}^{n} \gamma_{ij} f_{j} \tag{64}
\]

from which

\[
|\gamma_{ii} - 1| f_{i} > \sum_{j=1, j \neq i}^{n} \gamma_{ij} f_{j} = \sum_{j=1, j \neq i}^{n} \gamma_{ij} f_{j} \tag{65}
\]

because all \(\gamma_{ij}\) values are strictly between 0 and 1. Comparing (65) with (63) and ignoring the mass, as the same mass appears in every row \(i\), we see that matrix \(A\) is strictly diagonal dominant.

Furthermore, as \(\gamma_{ii}\) is strictly between 0 and 1, we have that all the diagonal elements of \(A\) are strictly negative. By Gershgorin circle theorem, all the eigenvalues of matrix \(A\) are strictly negative, and therefore, the matrix is Hurwitz. \(\square\)

**ACKNOWLEDGMENT**

The authors would like to thank IBM Zurich Research Laboratory for supplying measurement data of a real-life data center. This paper was carried out as part of the perspective program—Robust Design of Cyber-Physical Systems and Cooperative Networked Systems.

**REFERENCES**


Claudio De Persis received the Ph.D. degree in

system engineering from the Sapienza University of

Rome, Rome, Italy, in 2000. He was with the Department of Mechanical Automation and Mechatronics, University of Twente, Enschede, The Netherlands, and the Depart-

ment of Computer, Control, and Management Engi-

neering, Sapienza University of Rome. He was a

Research Associate with the Department of Systems

Science and Mathematics, Washington University,

St. Louis, MO, USA, from 2000 to 2001, and the

Department of Electrical Engineering, Yale University, New Haven, CT,

USA, from 2001 to 2002. He is currently a Professor with the Faculty of

Sciences and Mathematics, Engineering and Technology Institute, University of

Groningen, Groningen, The Netherlands. He is also with the Jan Willems

Center for Systems and Control. His current research interests include control

theory, and dynamical networks, cyber-physical systems, smart grids, and

resilient control.

Dr. De Persis was an Editor of the International Journal of Robust and

Nonlinear Control from 2006 to 2013, an Associate Editor of the IEEE

Transactions on Control Systems Technology from 2010 to 2015) and the IEEE Transactions on Automatic Control from 2012 to 2015.

He has been an Associate Editor of Automatica since 2013 and the IEEE


Pietro Tesi received the Ph.D. degree in computer

and control engineering from the University of

Florence, Florence, Italy, in 2010. He was a Research Associate with the University of Genoa, Genoa, Italy, from 2011 to 2013, and a

Visiting Research Scholar with the University of California at Santa Barbara, Santa Barbara, CA, USA, from 2008 to 2009. He is currently an Assist-
nant Professor with the University of Florence. He holds a guest research position at the University of Groningen, Groningen, The Netherlands. His current research interests include adaptive and learning systems, hybrid systems, and

networked control.

Dr. Tesi has been an Associate Editor of the IEEE Control Systems Society

Conference Editorial Board since 2014 and an Editor of the International Journal of Robust and Nonlinear Control since 2016.