Towards Energy Smart Data Centers: Simulation of Server Room Cooling System


Abstract— Cooling is an extremely important process in modern data centers. Cooling systems of server rooms ensure appropriate operation conditions to IT systems, such as servers and data storage, but, on the other side, they consume a lot of energy. Current control systems, which are installed in data centers and are responsible for thermal management of the facilities, are following conservative control strategies that maintain constant thermal conditions irrespective of computer load and outside temperature, thus efficient use of energy has not been appropriately addressed by them. In this paper, a method of optimizing energy consumption while maintaining an acceptable level of thermal comfort for CPUs in the server room is proposed. In the proposed method, a behavioral thermal model for server room should be created first, and then thermal behavior of the server room would be simulated under different circumstances using its thermal model, in order to find an optimum control strategy capable of retaining balance between thermal comfort and efficient use of energy. The effectiveness of the proposed method has been investigated via simulating a typical server room using MATLAB and SIMULINK and the results are demonstrated.

Keywords— data center automation; server rooms; thermal model; cooling system; energy efficiency;

I. INTRODUCTION

The number of data centers has undergone a rapid growth as data processing has had an increasing demand during the past two decades, causing fast increment of energy consumption in this sector. For instance, according to [1] more than 2% of the US total electricity usage is consumed by data centers out of which 50% is used for cooling server rooms. This amount is reported to be 1% in the world and it is expected to grow up to 2% in the near future and over 30% of this energy is used in cooling sector [2]. As a close observable site to the authors, Facebook data center in Northern Sweden consumes the comparable amount of energy (100MW) to the SSAB coking steel plant that previously has been the largest energy consumer in the region. The increasing amount of energy used for data center cooling systems is a warning sign and a strong motivation for the researchers to propose methods for reducing it.

In general, data centers comprise of 2 groups of equipment; CPUs and IT related equipment, which are the main source of producing heat, and the cooling devices responsible for removal of harmful heat [18]. In this research our main focus is the free airside cooling system, which is defined as “utilization of the natural climate to cool a server room” [1]. As described in [13] airside free cooling has 2 types 1) Direct: drawing the cold air directly into the data center, and consists of fans, dampers and controllers 2) Indirect: pumping the cold air into heat exchanger first. In this research, for the sake of simplicity we focus on the direct method. However, no matter what method is used; cooling is a highly energy consuming process. Therefore, efficient methods of energy and thermal management [14,17] must be utilized in order to reduce the unnecessary energy usage and wasted energy. Furthermore, modern data centers are very complex systems, thus efficient optimization methodologies [16] must be used to optimize the amount of energy, and achieve energy efficiency [15].

In this paper we use an optimization method, which has been described in [9,10,11,12], and it is based on modeling thermal behaviors of the cooling systems and then designing a controller that can control the cooling devices in an energy efficient manner. Our proposed method is based on simulating the thermal behavior of a typical cooling system consisting of fans and CPUs, using direct airside free cooling. The simulation is performed in a specific period of time under defined constraints and different conditions and calculates the energy consumptions and CPU temperatures for different scenarios separately. Next, we compare the results and select a scenario that consumes least energy while maintaining a minimum thermal comfort level for the CPUs. The results of this decision-making process then will be used to design an energy efficient control strategy for cooling system. The thermal modeling techniques and simulation are implemented in MATLAB/ SIMULINK. The result of the aforementioned simulations is intended to be used to enhance energy efficiency of the current Building Automation System (BAS) [4] of datacenters managing building’s heating, ventilation and air conditioning (HVAC) [3].

The rest of this paper is organized as follows: Section II is a brief background study on thermal modelling of cooling systems of server rooms. In Section III the implementation of thermal models in MATLAB/SIMULINK are described. Section IV describes the simulation scenarios and Section V demonstrates the simulation results. In section VI conclusion and some suggestions for future works are presented.

II. BACKGROUND STUDY

Energy efficiency of data centers depends on efficient energy consumption in server rooms. The major energy consumers in the server rooms are CPUs and cooling devices. Power consumption of CPUs is determined by their
computational load, while cooling systems typically operate based on thermal conditions and airflow [5]. For efficient use of energy, it is necessary to find a model that allows evaluating the cooling system in terms of energy consumption while maintaining a permissible temperature range. In this section power consumption and the thermal behaviors of the server room are modeled. The model includes power consumption of CPUs and cooling system.

Power consumption of computational nodes has significant impact on its temperature and CPU is a major energy consumer amongst computational nodes, thus power consumption of other parts are ignored in this preliminary research. However, the CPU power usage depends on its utilization (U), measured as a percentage of total CPU capacity and can be estimated with equation (1) adapted from [5, 6]:

\[ P_{CPU} = P_{idle} + \left( P_{max} - P_{idle} \right) U \]  

Equation (1) is valid under the assumption of known value of power consumption (W) and thermal characteristics in the model. Equation (6) adapted from [10] allows calculating the rate of CPU temperature change (\( T_{CPU} \)):

\[ T_{CPU} = \frac{P_{CPU}}{C_{CPU}} \left( T_{CPU} - T_{in} \right) \]  

Here \( P_{CPU} \) is CPU’s power consumption; \( T_{CPU} \) is the CPU’s temperature and \( T_{in} \) is the temperature of the air at the inlet of the CPU. \( C_{CPU} \) is the heat capacity of CPU and \( R \) is the CPU thermal resistance.

The thermal temperature of the air at the outlet of the CPU (\( T_{out} \)) is necessary for calculating the heat exchange rate (\( Q \)) between CPU and air. These values are calculated by equations (7) and (8) adapted from [9]:

\[ T_{out} = \left( 1 - \frac{1}{c_p \cdot f_{kg}} \cdot \frac{1}{R} \right) T_{in} + \frac{1}{c_p \cdot f_{kg}} \cdot \frac{1}{R} \cdot T_{CPU} \]  

Here \( c_p \) is the specific thermal capacity of the air; \( f_{kg} \) is rate at which air flows over the CPU, measured as kg/s.

Equation (8) is suitable for calculation the rate of heat exchange provided by GF. \( T_{out} \) is the ambient temperature and \( T_{in} \) is the temperature in the server room.

\[ Q(t_0 + \Delta t) = c_p \cdot m \cdot T_0 + \sum_{i=1}^{n} \int_{t_0}^{t_0 + \Delta t} Q_{CPU,i}(t) dt \]  

For describing the server room’s thermal evolution equations (9-10) are used adapted from [10].

\[ T_{SR}(t_0 + \Delta t) = \frac{Q(t_0 + \Delta t)}{c_p \cdot m} \]  

III. MATLAB IMPLEMENTATION OF MODEL

The implementation of model in MATLAB contains a set of system’s parameters and a collection of functions for calculating base values. TABLE I shows the set of parameters and TABLE II illustrates MATLAB implementation of equations described in previous section. The explanation about
each parameters and functions are written as comments in TABLE I AND II.

### TABLE I. SET OF PARAMETERS IN MODEL MATLAB IMPLEMENTATION

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p_{idle}$</td>
<td>50 W</td>
</tr>
<tr>
<td>$P_{CPU}$ power consumption in its idle state (W)</td>
<td>115 W</td>
</tr>
<tr>
<td>$T_{SR}$</td>
<td>21 °C</td>
</tr>
<tr>
<td>$T_{max}$</td>
<td>70 °C</td>
</tr>
<tr>
<td>$c_{cpu}$</td>
<td>30 J/kg°C</td>
</tr>
</tbody>
</table>

### TABLE II. MATLAB IMPLEMENTATION OF MODEL EQUATIONS

```matlab
%calculation of CPU power consumption
function [p] = cpu_p_calc(U)
    p = p_idle + (p_max - p_idle) * U / 100;
end
%calculation of LF power consumption
function [p] = lf_p_calc(rpm)
    p = p_max * (rpm / p_max) / 3;
end
%calculation of GF power consumption
function [p] = gf_p_calc(f_CFM)
    p = p_max * (f_CFM / 9); %calculation of GF power consumption
end
```

IV. SIMULATION SCENARIOS

This section focuses on the simulation scenarios. As shown in Fig. 1, the simulation configuration consists of 10 identical CPUs divided equally into 2 racks, 10 identical LFs and 1 GF. Control system of this configuration consists of two IEC 61499 CPUs, which were used for the calculation of heat amount produced by the system.

Fig. 1. Simulation configuration.

The time period of simulation is 1 hour (3600 seconds), which is divided into 4 equal parts each one which is 15 minutes (900 seconds) and all CPUs work in the following way: 1) in the first quarter the utilization of each CPU, $U = 20\%$, 2) in the second quarter $U = 60\%$, 3) in the third quarter $U = 40\%$, 4) in the fourth quarter $U = 80\%$. The goal of this simulation is to choose operation mode of fans supporting permissible values of CPUs’ temperature while providing energy savings. The local fan’s speed and duration of its operation are configurable parameters of the desired mode. Moreover, we not only would like to find the lowest speed for the fans, but also we would like to find out if in any case we can completely turn-off the fans for some period of time, to save even more energy.

V. SIMULATION RESULTS

Since LFs can operate in 3 speeds, the first or the initial scenario was conducted to find out which speed is the most energy efficient. Thus it consists of 3 cases each one of which is corresponded to one of the LF’s speeds (low, medium or fast). In all cases global fan worked only 2 minutes at the beginning of each quarter. Values of base the parameters of the model are shown in TABLE III for all 3 cases. In this table each row presents a quarter of the simulation time. As the table suggests, total CPUs temperature at the end of the simulation are 50.9, 47.2 and 46 consequentially for low, medium and fast LF speeds and none of them exceeds the max CPU temperature of 70°C. Although the low speed causes highest CPU temp, we choose it since it has lowest energy consumption of 1803.3 W and therefore, saves energy. Fig.2 illustrates the values of CPU’s and server room’s temperature corresponding to data presented in TABLE III.

The second scenario has been performed to optimize the results of the initial scenario by reducing LFs operation time.
under following conditions: each quarter of modeling time was divided into two periods; the first period: when all LFs are idle, with time duration of $t_{idle}$ and the second period when LFs work with time duration of $t_{op}$ and constant speed (three cases: low, medium and fast). The goal of scenario 2 therefore is to find appropriate values for $t_{idle}$ and $t_{op}$ so that CPUs temperatures will not exceed the max value. For this purpose we start from $t_{idle} = 8 \text{ min}$, $t_{op} = 7 \text{ min}$, which is approximately in the middle of a quarter. Fig. 3 illustrates that the CPUs temperatures exceed the max value with this configuration for all 3 fan speeds. Therefore, the simulation tool will randomly reduce the value of $t_{idle}$ until an idle time for LF is found which does not cause overheating of the CPUs.

The third scenario is about randomly changing the idle time for LF and finding a stable value for it. As illustrated in Fig. 4 if we select the following values: $t_{idle} = 2 \text{ min}$, $t_{op} = 13 \text{ min}$. CPUs temperatures do not exceed the max value which is indicated in this figure by red color, for none of the LF speeds and therefore all of these speeds are suitable to provide the appropriate comfort level to CPUs. TABLE IV presents values of the base parameters for this simulation scenario. Since slow LF speed saves the energy and makes no harm to the CPUs, it is justified to use low LF speed for comparing the total energy consumption of scenario 1 and 3. From TABLE IV Total Energy consumption is 1798.2 Wh for scenario 2 while from TABLE III this amount is 1803.3 Wh for scenario 3. Therefore we can conclude that reducing the LF operation time could save more energy (around 5 extra W per 1 hour).

Another factor that can be examined to save even more amount of energy is reduction of GF operating time, and therefore scenario 4 focuses on this aspect. Let GF operates only 1 minute (instead of 2 minutes) in the beginning of each quarter of modeling period. Fig. 5 shows that for this GF operating mode and for LF $t_{idle} = 1 \text{ min}$, $t_{op} = 14 \text{ min}$ the values of CPUs temperatures do not reach the max value which is indicated in this figure by red color, for none of the LF speeds. TABLE V presents values of base parameters of the model. Comparing the total energy consumption for low LF speed from TABLE IV for Scenario 3 (1798.2 Wh) and from TABLE V for scenario 4 (1321.4 Wh) indicates even a higher energy saving (around 477 Wh). However, as Fig. 5 suggests, reducing GF operation time is not good for long term planning, as it will cause continuous increment in server room temperature, which in some point of time will exceed the max server room temperature.

![Fig. 2. The values of the CPU and server room temperature at the continuous operation of LFs](image)

TABLE III. Values of base parameters of model in case of continuous operation of local fans with constant speed and operation of global fan during 2 minutes at the beginning of each quarter of modeling period

<table>
<thead>
<tr>
<th>Time, s</th>
<th>Server Room</th>
<th>Temperature, °C</th>
<th>Energy consumption, Wh</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CPU Temp for Different LF Speeds</td>
<td>CPUs</td>
</tr>
<tr>
<td>900</td>
<td>20.6</td>
<td>38.5</td>
<td>36.2</td>
</tr>
<tr>
<td>1800</td>
<td>21.1</td>
<td>46.4</td>
<td>43.2</td>
</tr>
<tr>
<td>2700</td>
<td>21.1</td>
<td>42.7</td>
<td>40.0</td>
</tr>
<tr>
<td>3600</td>
<td>21.9</td>
<td>50.9</td>
<td>47.2</td>
</tr>
</tbody>
</table>

TABLE IV. Values of base parameters of model in discontinuous operation of local fans with constant speed and operation of global fan during 2 minutes at the beginning of each quarter of modelling period

<table>
<thead>
<tr>
<th>Time, s</th>
<th>Server Room</th>
<th>Temperature, °C</th>
<th>Energy consumption, Wh</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CPU Temp for Different LF Speeds</td>
<td>CPUs</td>
</tr>
<tr>
<td>900</td>
<td>20.6</td>
<td>38.5</td>
<td>36.2</td>
</tr>
<tr>
<td>1800</td>
<td>21.1</td>
<td>46.4</td>
<td>43.1</td>
</tr>
<tr>
<td>2700</td>
<td>21.1</td>
<td>42.7</td>
<td>40.0</td>
</tr>
<tr>
<td>3600</td>
<td>21.9</td>
<td>50.9</td>
<td>47.2</td>
</tr>
</tbody>
</table>

### TABLE V.
VALUES OF BASE PARAMETERS OF MODEL IN DISCONTINUOUS OPERATION OF LFS ($t_{idle} = 1\text{ min}$, $t_{op} = 14\text{ min}$) AND OPERATION OF GLOBAL FAN DURING 1 MINUTE AT THE BEGINNING OF EACH QUARTER OF MODELING PERIOD

<table>
<thead>
<tr>
<th>Time, s</th>
<th>Server Room</th>
<th>CPU Temp for Different LF Speeds</th>
<th>Energy consumption, Wh</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>CPUs Slow</td>
<td>Med</td>
</tr>
<tr>
<td>900</td>
<td>21.7</td>
<td>39.6</td>
<td>37.3</td>
</tr>
<tr>
<td>1800</td>
<td>23.2</td>
<td>48.5</td>
<td>45.3</td>
</tr>
<tr>
<td>2700</td>
<td>24.2</td>
<td>45.8</td>
<td>43.0</td>
</tr>
<tr>
<td>3600</td>
<td>25.9</td>
<td>54.9</td>
<td>51.2</td>
</tr>
</tbody>
</table>

**Fig. 3.** The values of the CPU and server room temperature at the discontinuous operation of LFs ($t_{idle} = 8\text{ min}$, $t_{op} = 7\text{ min}$)

**Fig. 4.** The values of the CPU and server room temperature at the discontinuous operation of LFs ($t_{idle} = 2\text{ min}$, $t_{op} = 13\text{ min}$)

**Fig. 5.** The values of the CPU and server room temperature when GF operating 1 minute at the beginning of each quarter of simulation period and discontinuous operation of LFs ($t_{idle} = 1\text{ min}$, $t_{op} = 14\text{ min}$)

### TABLE VI.
COMPARISON OF SCENARIOS

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Total Energy Consumption,(Wh)</th>
<th>CPU temperature</th>
<th>Energy Reduction</th>
<th>Percentage of Saving</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base - Scenario</td>
<td>2196.1</td>
<td>Less than 70°C</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Scenario1</td>
<td>1803.3</td>
<td>Less than 70°C</td>
<td>393.1</td>
<td>17.9 %</td>
</tr>
<tr>
<td>Scenario2</td>
<td>1792.1</td>
<td>Exceeds 70°C</td>
<td>Not Acceptable</td>
<td>Not Acceptable</td>
</tr>
<tr>
<td>Scenario3</td>
<td>1798.2</td>
<td>Less than 70°C</td>
<td>397.9</td>
<td>18.1 %</td>
</tr>
<tr>
<td>Scenario4</td>
<td>1321.4</td>
<td>Less than 70°C</td>
<td>874</td>
<td>39.8 %</td>
</tr>
</tbody>
</table>

The total energy consumptions of all 4 scenarios as well as a base scenario are shown in TABLE VI. The Base-Scenario is the default cooling strategy which is based on maximum (fast) LF speed, full operation time of LFs and 2 minutes of GF.
The energy consumption of the Base-Scenario is taken from Table III for fast LF speed. In addition Scenario 2 is not acceptable because it cannot maintain the CPU temperature below max limit. In the first iteration, the proposed method could reduce the energy consumption of the Base Scenario to 393.1 and save 17.9 % of the total Energy consumption. Scenario 2 is not acceptable as mentioned earlier. Scenario 3 shows 397.9 Wh energy reduction and 18.1 % saving. Finally Scenario 4 reduces the total energy consumption by 874 Wh and saves 39.8 % of the total energy, which has been consumed in Base-Scenario. In conclusion, the results show that considerable amount of energy can be saved by applying the proposed method.

VI. CONCLUSION AND FUTURE WORK

In this paper a simulation-based optimization method for energy consumption of cooling system for a typical server room consisting of 2 racks, 10 CPUs, 10 local fans and a global fan is proposed. The method consists of thermal modeling of the server room and its components, implemented in MATLAB and SIMULINK. To examine the effectiveness of the method, 4 simulation scenarios have been defined and conducted. The main objective of these scenarios is to find out the possibility of reducing energy consumption of the cooling devices, while maintaining acceptable temperature for the CPUs below their maximum allowed temperature limit. By simulating the cooling system under different conditions such as fan speeds and operation times, the proposed method could considerably reduce the energy usage while maintaining the thermal comfort for CPUs. The thermal model, simulation scenarios and results have been described and demonstrated as well. As future work to this research, we are planning to create a complete physical model for the server room as well as control system for controlling the cooling system, and then simulate more complex situations using hardware in the loop approach.

ACKNOWLEDGEMENT

This paper is an output of the CEED Project with Project Number: 39357-1, which is partially funded by Energymyndighetten (Swedish Energy Agency) and Luleå University of Technology and funded in kind by ABB Cooperate Research in Västerås Sweden.

REFERENCE