THERMAL ANALYSIS OF A DATA CENTRE COOLING SYSTEM UNDER FAULT CONDITIONS

Michaël Kummert[1], William Dempster[1], and Ken McLean[2]

[1]Mechanical Engineering University of Strathclyde, Glasgow, Scotland
[2]WSP UK Ltd, Glasgow

ABSTRACT

This paper describes the modelling and analysis of a 5 MW chilled water plant used for the cooling of a recently constructed data centre in the UK. The model was developed using the TRNSYS software with the aim of studying the impact of perturbations such as chiller failure on the water and air temperatures in the system. The model includes the chillers, hydraulic network (piping, valves and pumps) and individual water-to-air heat exchange units. The coupling between the data centre air temperature levels and the cooling plant has enabled a full assessment of the cooling system design in response to system perturbations to be carried out. The paper examines a number of scenarios involving the failure of the chillers and shows how the inherent thermal inertia of the system plus additional inertia achieved through buffer vessels allowed a suitable design to be achieved. The effective sizing of these vessels could not have been achieved with confidence without the use of transient analysis of the complex system dynamics.

INTRODUCTION

The growth of data centre demand in the last 10 years has led to an increase in their size and reliability requirements with a significant effort given to ensure operational effectiveness. For example, the largest data centre recently commissioned being the Microsoft Chicago data centre requiring nearly 200 MW of power, 500,000 ft^2 and costing \$500 million (Reuters). Data centres housing the IT infrastructure of large organisations constitute a considerable technical challenge to ensure even the minimum requirements of 96.67% reliability for the lowest Tier 1 operational availability. For mission critical IT systems, the Tier 4 criteria require significant redundancy and therefore cost to ensure the 99.995% reliability (Turner et.al). One of the major requirements of continuous operation is the cooling of the IT systems. The IT infrastructure accounts for some 50% of the data centre power needs; the remaining being dominated by the electrical power system and the cooling equipment. Mission-critical data centres such as the one considered in this study require tight environmental control that meet the "class 1" requirements defined by ASHRAE (ASHRAE, 2004).

Specifying cooling systems to maintain suitable temperature levels and dissipate the heat generated can be carried out using industry standard design methods. However, accounting for perturbations in cooling due to failure of plant and restart of backup systems requires for faster thermal transients to be addressed than would normally be encountered in building system analysis. The most common approach to the design of large data centres is to cool the IT systems with fan driven air cooling , the heat from which is transferred via air-water heat exchanger units to chilled water circulating in a central ring main. Water chillers are then used to extract the heat and transfer this to the atmosphere. To ensure reliable operation and address failure in the major cooling components redundancy is built into the system (see e.g. Hasan, 2009). The issue considered in this paper are the consequences of a failure in the chiller systems and the inherent time delays that result when starting a chiller or when changing load. The consequences are that the cooling system may not meet the cooling requirements temporarily and thus rely on the thermal inertia of the chilled water existing in the pipework to keep the IT system within their operation limits. The transient response of the system is partly determined by the hydraulic circuit transit times and since this is of the order of minutes it imposes a level of analysis that is not normally required in building system simulation.

Furthermore, to address the impact on the IT cooling, the thermal coupling between the server data rooms, the chilled water circuit and the chillers all need to be accounted for. This results in a model that encompasses the whole cooling system with the smallest scale being that of the individual heat exchanger unit. This paper examines these issues by presenting a study carried out during the design phase of a tier 3 data centre designed by WSP UK Ltd on behalf of Norwich Union (now Aviva). Of particular interest was the effect of the loss of cooling caused by chiller failure and the resulting temperature transient that propagated to the IT halls. This paper describes the model used to analyse the cooling system and the failure scenarios that were considered. The transient response of system temperatures (chilled water and data room air temperature) are presented for two of the chiller failure scenarios. The effect of including added thermal inertia into the system by including buffer vessels is also examined

DATA CENTRE AND MODELLING

Data Centre Cooling System

The cooling system for the data centre is shown in Figure 1. The heat load comes primarily from the IT requirements with a small percentage from office air conditioning. The IT load is distributed between 46 Room Air Conditioning Units (RACUs), and 24 water-based cooling units referred to as Heat Exchanger Pumping Units (HEPUs), which are supplied by cooling water from a chilled water pipework system. The pipework consists of a 300 mm diameter supply ring main and a return ring main adding to a total length of over 1500 m of pipework. Each ring main has two inlet and two exit connections which delivers and removes water from each half of the rings providing a flow split which allows both clockwise and anti-clockwise circulation of the cooling flow. The flow in the supply ring is continuously extracted for RACU use from the supply pipework, which results in a natural depletion of the flow by RACU 28, see Figure 1. The IT room cooling is achieved with an air circulation system as shown in Figure 2. Cold air is supplied via floor vents and hot air extracted via ceiling vents with the RACU's extracting the heat to the chilled water using air-water heat exchangers. Two optional buffer tanks each consisting of approximately 30% of the system mass and used to increase the thermal inertia of the system are also shown in Figure 1. Their effect on the system response will be discussed later. The circuit flow is provided by up to 6 parallel primary pumps and the heat load is removed using up to 6 parallel chillers.

.

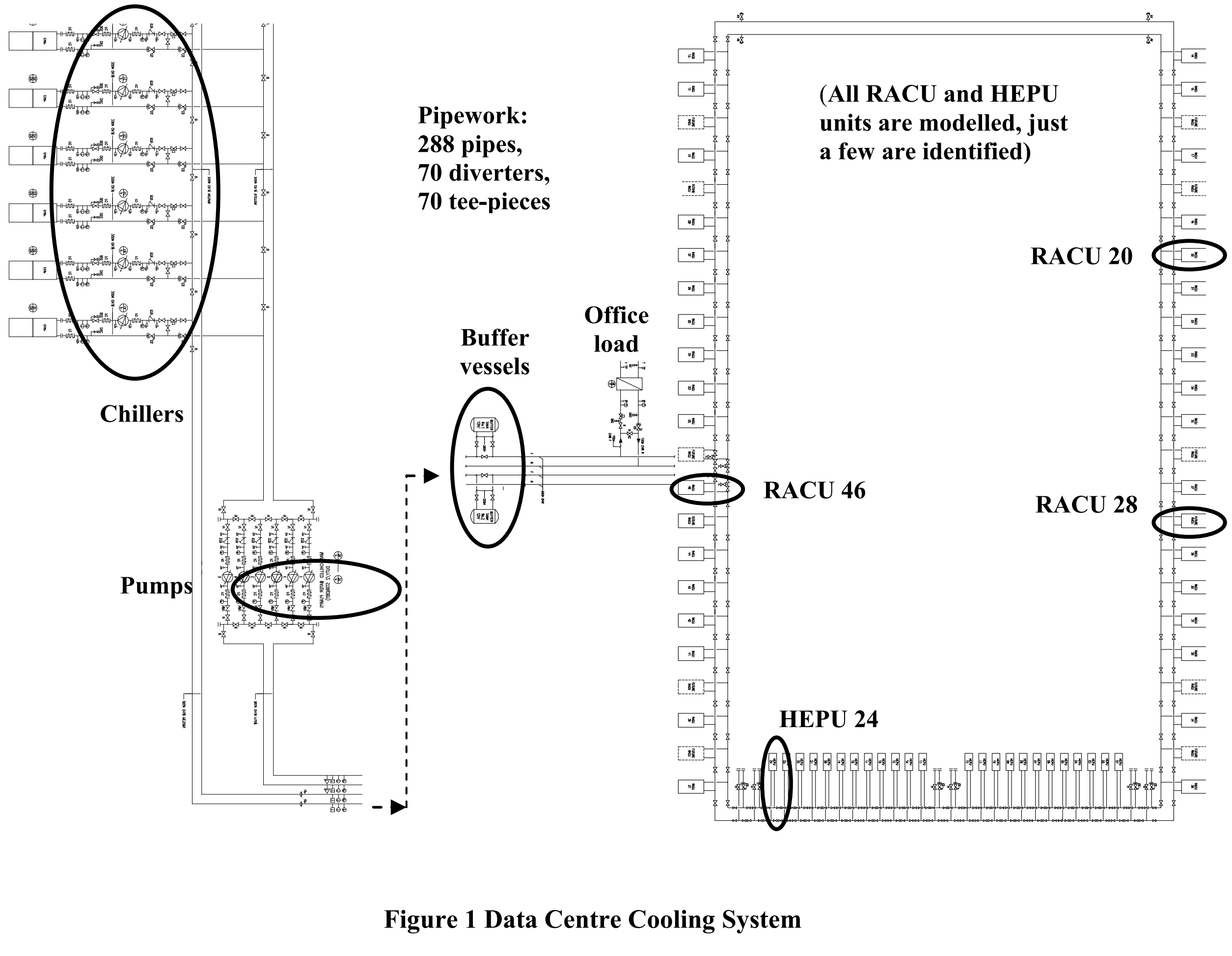

Figure 1 Data Centre Cooling System

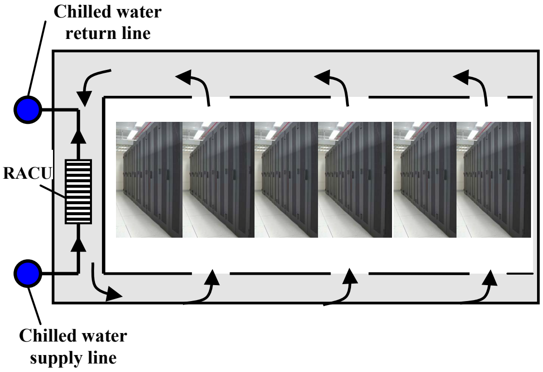

Figure 2 IT Hall cooling system

The total design heat load for the system is 4.6 MW with a control system that requires to maintain the IT room air conditions below 24 °C for chilled water supplied at 9 °C. The redundancy that is required of this Tier 3 data centre consists primarily of additional chiller capacity and over specification of the heat transfer capacity of the RACU and HEPU units. During normal operation only 4 chillers are required with 2 being redundant. The RACU's are operated at approximately 85% of their capacity with 4 of the 6 pumps in normal operation. These features provide a total cooling capacity of 7.8 MW, which is 40% above the normal operating design loads. However, these requirements are not in themselves sufficient to ensure that the IT equipment will remain cooled under transient conditions resulting from plant failure. In this work the effect of failure in up to two chillers was examined; i.e. the required use of all of the redundant plant. Since the chillers considered for the Data Centre take up to 8 minutes to reach full capacity and the effect of a loss of cooling would have propagated to the RACU's within this time a number of scenarios were conceived to minimise the air temperature upsets in the IT halls of the data centre. This involved minimising the chiller start up time by running at part load more than the four chillers normally required for cooling. The time for a chiller to increase from a part load of 60% to a full load could be as short as 15% of the time to start up from cold. A number of different scenarios were examined but only two extremes discussed here to show the main effects, as follows

(i) 4 chillers running at 90% load , 2 awaiting start up ("N" configuration)

(ii) 6 chillers running at 60% load ("N+2")

Different buffer vessel sizes are also compared in the results presented below.

Plant Modelling Approach

The model has been developed and implemented in TRNSYS (Klein et al., 2006), a component-based integrated energy performance modelling tool, using a combination of existing component models and bespoke components. The model includes a simplified but complete description of the hydraulics in the chilled water system and the interaction with the air-side cooling systems in the IT hall. This enables assessing the impact of chilled water temperature disturbances on the overall IT environment. The hydraulic network shown in Figure 1 is modelled faithfully using standard TRNSYS components for pumps, diverters and tee-pieces. Each RACU is modelled separately with an individual TRNSYS component. HEPU's only handle 17% of the load and since their specification was unknown these have been modelled as RACU's. However, it has been assumed that the thermal inertia in the heat exchangers is low compared to the thermal mass of the rest of the system which allows standard heat exchanger theory to be applied. The standard pipe component in TRNSYS accounts for travel time and for the thermal mass of the fluid in the pipe but not for the thermal mass of the pipe wall and insulation. A more detailed pipe model was obtained (TESS, 2007) in order to take full account of the thermal inertia provided by the mass of steel pipework. This thermal mass has a notable impact on the dynamic system response.

Room-side heat transfer is modelled by splitting the IT hall into different "zones". Each zone represents the notional part of the hall served by each RACU unit. The Zone air volume and floor area are proportional to the share of the total room-side cooling load provided by the corresponding RACU. A schematic of a zone and its associated RACU is represented in Figure 3. No interaction is assumed between the different zones.

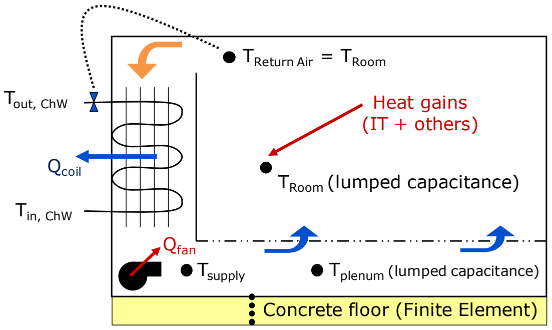

Figure 3 IT hall air heat transfer model

Two "air nodes" (air volumes assumed to be at a uniform temperature) are modelled for each zone: the IT hall and the ventilation plenum under the raised floor. Thermal mass is added to the nodes to account for IT equipment and other structures, while the heavy concrete floor is modelled by a finite element approach. This simple model was implemented in a bespoke TRNSYS component. The calculated return air temperature only provides an approximate indication of the air temperature surrounding the IT equipment (a full CFD model would otherwise be required) but the model is accurate enough to represent the room response to changes in chilled water supply temperature and its impact on the chilled water flowrate and exhaust temperature. The distribution of the heat gains are presented in Table 1 and it can be seen that the majority of the heating is from the IT Hall, i.e. 70% of the total load with the remainder from the cooling equipment, office loads or gains from the surrounding atmosphere via pipe walls.

Description	Cooling load	
IT cooling load	3375	kW
Other room loads	215	kW
RACUs fan gains	390	kW
HEPUs fan and pump gains	375	kW
Office air conditioning	200	kW
Piping, pumps	45	kW
Total cooling load	**4600**	**kW**

Table 1 Data Centre heat gains

The IT room cooling loads are distributed evenly between the RACU's. The design operating conditions for these heat loads are to maintain the IT data room below a maximum temperature of 24 °C. As mentioned above the RACUs are operating well below their rated capacity, which results in a flowrate equal to 50 % of the rated value and a chilled water return temperature of 19.4 °C (supply temperature = 9 °C).

Control dynamics

The plant cooling systems have a number of dynamic constraints that can have an impact on the response of the system. The most significant is the start up of the standby chillers or if running the change in load response. This is shown in Figure 4, which indicates that the chiller produces no cooling for the initial two minutes after starting and then ramps to full load over a period of 5 minutes. It is the impact of this period of reduced cooling that is studied in this paper.

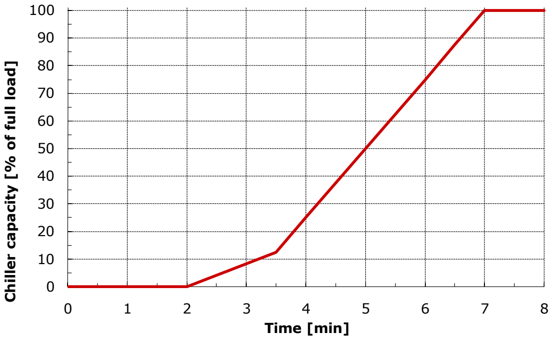

Figure 4 Chiller load-time response

The flowrate to the chillers is controlled by butterfly valves that have a 20-second response time for a fully open to a fully shut condition or vice versa. This relative short time period can have a significant effect on the initial transient due to the nonlinear flow-valve area characteristic, which is shown on Figure 5. This characteristic indicates a fast opening but a slow closure, which means that during the initial part of a transient all 6 chillers will be supplied with a water flow. However, only 2 chillers may be operating in a cooling mode immediately after a failure thus only cooling one third of the flow. The consequences on the cooling water temperature will be seen shortly.

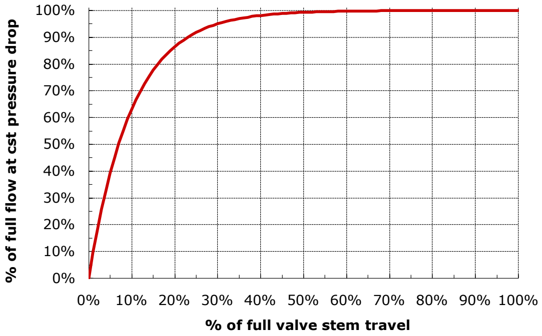

Figure 5 Chiller Control Valve Flow- Area Curve

There are further control issues that have been modelled but have a lesser effect. When a failure is detected, the BMS (Building Management System) resets the chilled water setpoint to 6 °C in order to utilise all the available chiller capacity. The flowrate in each RACU is controlled by a 2-way valve while secondary pumps are controlled to maintain a constant pressure difference across selected RACUs. The 2-way valves are controlled by local controllers (PID, Proportional Integral Derivative) in order to maintain a constant return air temperature at 24 °C. The response of these PID controllers has an impact on the overall transient system response and different assumptions were compared. The "average" assumption was kept for the results presented in this paper.

The behaviour of system components (e.g. curves presented in Fig. 4 and 5) was derived from information available from the manufacturer. In the context of this project, it was not possible to compare them to actual performance data measured by the Building Energy Management System.

CALCULATION RESULTS

System Response to Scenario (i)

The worst-case scenario is when 4 chillers are operating, two fail and the redundant chillers start up from cold. The system response to this cooling failure is taken as the reference case and used to show the main issues that arise when a chiller failure occurs. Figure 7 shows the chiller response with the total cooling being substantially reduced at time zero due to the failure of chillers 3 and 4. Chillers 1 and 2 see a reduced flowrate for about 20 seconds because all the butterfly valves open at the same time. After that, they start to ramp-up to their full capacity, responding to the "failure mode" chilled water setpoint of 6 °C instead of 9 °C. Chillers 1 and 2 reach their full capacity after less than one minute. During this time chillers 5 and 6 are still not producing cold. They only start to ramp-up after 3 minutes and they reach their full capacity after 8 minutes.

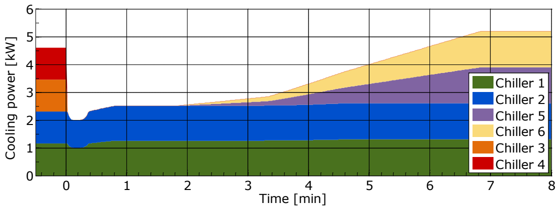

Figure 7 Chiller response during failure of Chillers 3/4 and start of chiller 5/6

The reduction in cooling along with delayed opening and closing of the chiller control valves (Figure 5) produces an increase of 7 °C in the overall chilled water temperature as can be seen in Figure 8. The temperature remains warmer until the cooling capacity of the chillers recovers. This surge of hotter water propagates through the cooling system as can be seen in Figure 9 where the cooling water is identified at different locations through the circuit, (refer to figure 1 for locations).

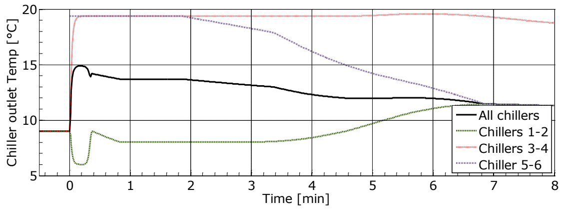

Figure 8 Chiller water Outlet temperature

This temperature change in the cooling water leads to a reduction of over 40% in the water-air temperature difference across the HEPU and RACU's and if left unchecked would lead to substantial increases in the air temperatures of the data room. There are two features of the system which prevent this. Firstly, the rise in water temperature is partially reduced by heat transfer to the circuit pipework and can be seen in the reduction of the peak temperature as the hot water surge travels around the circuit, Figure 9.

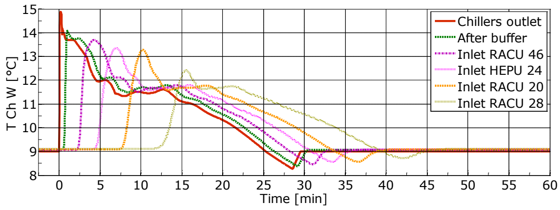

Figure 9 Chilled water supply temperature at different locations throughout circuit.

Secondly, the RACU and HEPU inlet valves are opened in response to an increase in air temperature resulting in an increase in flowrate, which is shown in Figure 10. In that figure, flowrate is expressed in % of the rated value, and as explained above the flowrate in normal operating conditions is around 50% of that value because of RACUs oversizing.

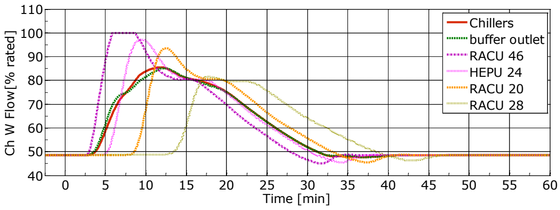

Figure 10 Chilled water flow at different locations

Thus, the effect of system thermal inertia combined with a control feature limits the temperature excursion and improves the heat transfer limiting the rise in IT hall air temperature to less than 1 °C at the first RACU, as shown in Figure 11.

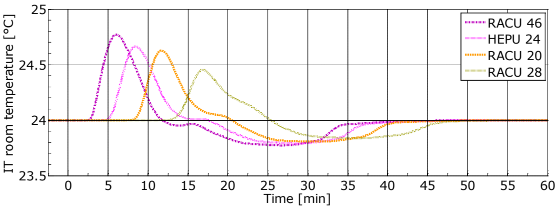

Figure 11 Room air (return) temperature

System Response to Scenario (ii)

The system response to chiller failure with 6 chillers running at part load was examined in Scenario (ii). Again two chillers are assumed to fail but since the remaining chillers are already running the time to increase from part load to full load is shorter than starting from cold and results in a much lesser transient. Figure 12 and 13 shows the overall cooling and the resulting air temperature increase. Figure 12 shows the impact on cooling is much less with only a 35% drop in capacity (see Figure 7) and the recovery time is shorter.

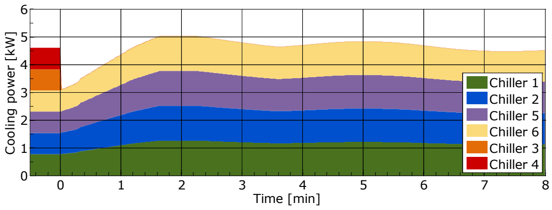

Figure 12 Chiller capacity during Scenario (ii)

The effect on air temperature is also significantly reduced, as shown on Figure 13, which uses the same scale as Figure 11.

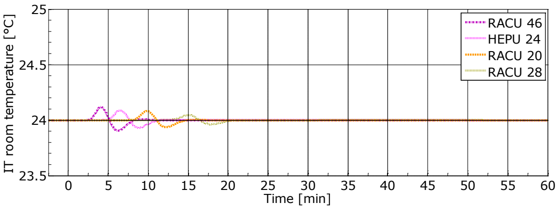

Figure 13 Room air temperatures, scenario (ii)

Effect of Buffer vessels

To improve the operating safety margin the thermal inertia of the system can be increased by the addition of buffer vessels. Due to the mixing of the inlet flow and the contents, the buffer vessel can only provide a reduction in the temperature excursion rather than negate it completely. Also since the vessel is of finite size the reduction can only be provided for a limited period. Thus the effect of adding a vessel and the sizing of the vessel was investigated. Vessels were added in both main inlet pipes downstream of the chillers, see Figure 1.

Two buffer vessel sizes were investigated: 2 x 6 m^3 and 2 x 12 m^3 and both conditions were run with the scenario (i) case, ie the worst-case chiller failure condition.

Figure 14 shows the effect of the buffer vessels on the chilled water temperature. Even without buffer vessels, the thermal mass in the (long) piping between the chiller plant room and the IT rooms introduces some damping, bringing the peak from 14.9 °C to 14.1 °C. The 2 x 6 m^3 vessel case shows a further reduction of the peak by 1.5 °C and the 2 x 12 m^3 vessel case a reduction of 0.9 °C. The initial temperature rise (9 to 14.1) is almost reduced by 50% in the case of the large vessel..

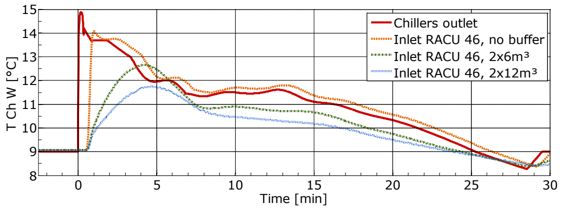

Figure 14 Chilled water temperature for scenario (i), different buffer cases

Figure 15 shows the effect of the buffers on the air temperature and shows a small reduction in the temperature increase by a about 0.5 °C, the air temperature largely being maintained by an increase in the liquid flowrate for each RACU or HEPU

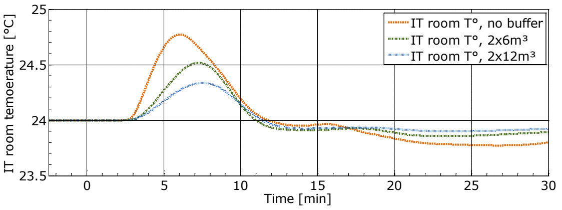

Figure 15 IT room temperature for scenario (i), different buffer cases

DISCUSSION

The results presented in this paper were obtained using a detailed transient model of a data centre chilled water cooling plant. The most significant simplifying assumption was made in modelling the IT rooms, which was split into "slices" served by each individual RACU. The lack of interaction between RACUs was retained to provide a worst-case estimate of the air temperature disturbances, given that individual RACUs will "see" chilled water supply temperature disturbances at different points in time. On the other hand, the simple model described in Figure 3 only gives a rough estimate of the conditions within the IT racks. A full CFD simulation would be required to investigate this.

The disturbances in chilled water supply temperature and in room air temperature may seem very small but they should be related to the strict environmental conditions imposed in this class 1 data centre. Recent work by ASHRAE has resulted in relaxed environmental criteria for data centres (ASHRAE, 2008), and servers performing well in extreme environments have been reported in the news (Miller 2008). But tier 3 and tier 4 data centres built today in the UK still have to meet very stringent criteria for environmental conditions and for the reliability and redundancy of cooling systems. This is mainly because the cost of service interruption is orders of magnitude higher than the cost of a few MW chilling capacity or the increase in energy cost that might result from very strict operating set points. The role of the detailed model described in this paper was to quantify the disturbances in the chilled water system temperatures and to design a system that would ensure that these disturbances would remain within the margins defined by the client. The model demonstrated that the system can be maintained within very tight operating air temperature margins through running redundant chillers (i.e. operating N+2 chillers) or adding relatively modest buffer vessels, as was verified for a failure scenario involving the simultaneous loss of two chillers.

In addition, the model provided valuable insight on the system behaviour in different operating conditions, e.g. by highlighting the very low chilled water velocities in the system resulting from a fully redundant distribution loop and highly part-loaded RACUs. The model is also useful to assess different scenarios such as running the RACUs at reduced fan speed or assessing the impact of controlling the supply air temperature rather than the return air temperature.

CONCLUSION

The response of a large data centre cooling system to failure in the chiller units is primarily dictated by thermal inertia of the system, the response time of the valve components and control strategies for the RACUs and HEPUs. The complex interaction that results requires a system modelling approach to identify the viability of the system to address operational upsets and to quantify the extent on the most important parameters. A detailed numerical model was developed in TRNSYS, using components taking into account the dynamic behaviour of all system components at the scale of one minute or less. The calculations carried out proved that when the system encountered sizable fluctuations in cooling water temperature that the designed system could maintain air temperatures in the IT hall to almost near normal operating levels. It is also beneficial that this could be achieved by several of the operating scenarios studied, whether by variations in chiller operation or by the increase in thermal inertia by the addition of buffer vessels.

REFERENCES

ASHRAE. 2004. Thermal Guidelines for Data Processing Environments. American Society of Heating, Refrigerating, and Air-Conditioning Engineers. Atlanta, GA, USA

ASHRAE, 2008. 2008 ASHRAE Environmental Guidelines for Datacom Equipment - Expanding the Recommended Environmental Envelope. American Society of Heating, Refrigerating, and Air-Conditioning Engineers. Atlanta, GA, USA

Hasan, Z., 2009. Redundancy for Data Centers ASHRAE Journal, 51(2), pp. 52-54.

Klein, S. A. et al., 2006. TRNSYS 16.1 – A TRaNsient SYstem Simulation program, User manual. University of Wisconsin-Madison, SolarEnergyLaboratory– http://sel.me.wisc.edu/trnsys

Miller, R., 2008. New from Microsoft: Data Centers In Tents. Data Center Knowledge website, http://tinyurl.com/4o7536, consulted in February 2009.

Reuters,2007. Microsoft invests $500 mln in Illinois data center. Reuters website, http://www.reuters.com/article/companyNewsAndPR/idUSN0757089820071108. Consulted in February 2009.

TESS, 2007. TRNSYS Type 604, multi-node pipe model. Personal communication. Thermal Energy Systems Specialists, Madison, WI.

Turner, W.P., Seader, J.H. Brill, K.G. Industry Standard Tier Classifications Define Site Infrastructure Performance, white Paper, The Uptime Institute, 2005.