THERMAL CHARACTERIZATION OF AN ISOLATED HYBRID COOLED SERVER WITH PUMP AND FAN FAILURE SCENARIOS

by

USCHAS CHOWDHURY

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree, of

MASTER OF SCIENCE IN MECHANICAL ENGINEERING

THE UNIVERSITY OF TEXAS AT ARLINGTON

MAY 2017

Acknowledgements

I would like to thank Dr. Dereje Agonafer who helped throughout my work. His continuous guidance and support over the last two years of my study and research at The University of Texas at Arlington helped me to cross all the obstacles.

I would like to thank Dr. Haji-Sheikh and Dr. Andrey Beyle for evaluating my work as committee members.

I would like to give a special thanks to Ms. Sally Thompson and Ms. Debi Barton for her immense help throughout my stay at UT Arlington

I would also like to thank all members at EMNSPC.

Finally, I would like to thank my parents who has always been the beacon of hope in my life.

I thank Almighty God for providing me the strength and inspiration.

May 21, 2017

Abstract

THERMAL CHARACTERIZATION OF A HYBRID COOLED SERVER WITH PUMP AND FAN FAILURE SCENARIOS

Uschas Chowdhury, MS

The University of Texas at Arlington, 2017

Supervising Professor: Dereje Agonafer

Modern day data centers are operated at high power for increased power density, maintenance, and cooling which covers almost 2 percent (70 billion kilowatt-hours) of the total energy consumption in the US. IT components and cooling system occupy the major portion of this energy consumption. Although data centers are designed to perform efficiently, cooling the high-density components is still a challenge. So, alternative methods to improve the cooling efficiency has become the drive to reduce the cooling cost. As liquid cooling is more efficient for high specific heat capacity, density, and thermal conductivity, hybrid cooling can offer the advantage of liquid cooling of high heat generating components in the traditional air-cooled servers. In this experiment, a 1U server is equipped with cold plate to cool the CPUs while the rest of the components are cooled by fans. In this study, predictive fan and pump failure analysis are performed which also helps to explore the options for redundancy and to reduce the cooling cost by improving cooling efficiency. The ASHRAE guidance class W4 for liquid cooling is chosen for experiment to operate in a range from 25°C - 45°C. The experiments are conducted separately for the pump and fan failure scenarios. Computational loads are applied while powering only one pump and the miniature dry cooler fans are controlled externally to maintain constant inlet temperature of the coolant. As the rest of components such as DIMMs & PCH are cooled by air, maximum utilization for memory is applied while reducing the number fans in each case for fan failure scenario. The components temperatures and power consumption are recorded in each case for performance analysis.

Table of Contents

Acknowledgements iii

Abstract iv

List of Illustrations vii

List of Tables viii

Chapter 1 Introduction 1

1.1 Data Center Energy Consumption 1

1.2 Thermal Management of Data Centers 2

1.5 Scope of Work 5

Chapter 2 Experiments 6

2.1 Server under Study 6

2.2 Test Setup 7

2.3 Test Procedure 10

Chapter 3 Results 14

3.1 Results for Pump Failure 14

3.2 Results for Fan Failure 16

Chapter 4 Transient Analysis 20

4.1 Full Pump Failure Scenario 20

4.2 Full Fan Failure Scenario 21

Chapter 5 Power Consumption 23

5.1 Power consumption in Pump Failure Scenario 23

5.2 Power consumption in Fan Failure Scenario 25

5.3 Reduction in Pumping Power 26

Chapter 6 Conclusion and Future work 27

6.1 Conclusions on Failure Scenarios 27

6.2 Future Work....28

References 29

Biographical Information....31

List of Illustrations

Figure 1-1 Typical thermal layout of a data center. 2

Figure 1-2 Environmental envelopes based on the class of data center 3

Figure 2-1 Cisco Server with retrofitted ASETEK Cold Plate 7

Figure 2-2 K-type Penetrable Thermocouple 7

Figure 2-3 Schematic of test setup and data acquisition 8

Figure 2-4 Experimental test setup and data acquisition 8

Figure 2-5 Front Panel of LabVIEW code 9

Figure 2-6 Pump failure scenario for downstream pump 11

Figure 2-7 Fan Failure scenario for 2 fans. 12

Figure 2-8 Fan Failure scenario for 3 fans 12

Figure 3-1 The variation in pump rpm for different voltages. 14

Figure 3-2 CPU Temp vs Coolant Temperature with 2 Pumps 15

Figure 3-3 CPU Temp vs Coolant Temperature with Single Pump 16

Figure 3-4 DIMMs temperature for 5,4,3 fan configurations at different fan PWM. 17

Figure 3-5 PCH Temperature for 5,4,3 fan configurations at different fan PWM. 17

Figure 3-6 Fan failure scenario with config 1..18

Figure 3-7 Fan failure scenario with config 2..18

Figure 3-8 Fan failure scenario with config 3..19

Figure 4-1 Full pump failure scenario with CPU temperatures vs time............................20

Figure 4-2 Full fan failure scenario with DIMMs and PCH temperatures vs time. 22

Figure 5-1 Power consumption at Maximum CPU Utilization with 2 pumps 23

Figure 5-2 Power consumption at Maximum CPU Utilization with 1 pumps 24

Figure 5-3 Power Consumption in Fan Failure Study 25

List of Tables

Table 5-1 Reduction in Pumping Power..26

Chapter 1

Introduction

To store, process and distribute large amounts of data, various IT equipment mainly servers, storage subsystems, networking switches, routers are housed in a facility called data center. The cabling and physical rack units are used to organize and interconnect the IT equipment. The facility also consists of power supplies, backup power supplies, environment controls (such as air-conditioning) and security measures.

1.1 Data Center Energy Consumption

The main purpose of Data center cooling technology is to run the IT equipment within suitable environment condition. The heat generated by IT equipment is continuously transferred to the environment or a sink while substantial amount of energy is used to maintain a certain operating temperature for the facility and IT components. Over the years, there has been an increment of power consumption by data centers. According to a report by US data centers consumed about 70 billion kilowatt-hours of electricity in 2014, the most recent year examined, representing 2 percent of the country's total energy consumption, according to the study [1]. The increase in energy usage is a concern but there is also a good news about the reduction in the trend of energy consumption by data centers while comparing the data for the last decade. More measures are needed to be taken to improve the data center cooling efficiency. This includes use of ambient air, warm water cooling, chiller less cooling etc.

The most common metric used to measure efficiency of data centers is the Power usage effectiveness (PUE) which was proposed by the Green Grid. PUE is a ratio of the total energy to the energy consumed by IT equipment. A PUE of 2.0 means that for every watt used by IT equipment, another watt is used by cooling, power distribution, and lighting equipment. Most operators consider a 1.65 PUE (the average for the survey) good enough, even though some hyper-scale cloud providers are consistently achieving PUEs below 1.1 in their data centers [2]. The global average PUE of data centers is between 1.8 and 1.89 [3].

1.2 Thermal Management of Data Centers

Computer based application requires faster and improved communication. To meet with the loads, efficient servers are necessary for processing. A substantial amount of energy is required to cool of the servers with high heat generating components. Thermal management of data center is multi-scaled. The different levels of thermal management which needs consideration are chip level, server level, rack level and room level. The primary heat generates at chip level and disperses into the data center which is cooled by the CRAC unit.

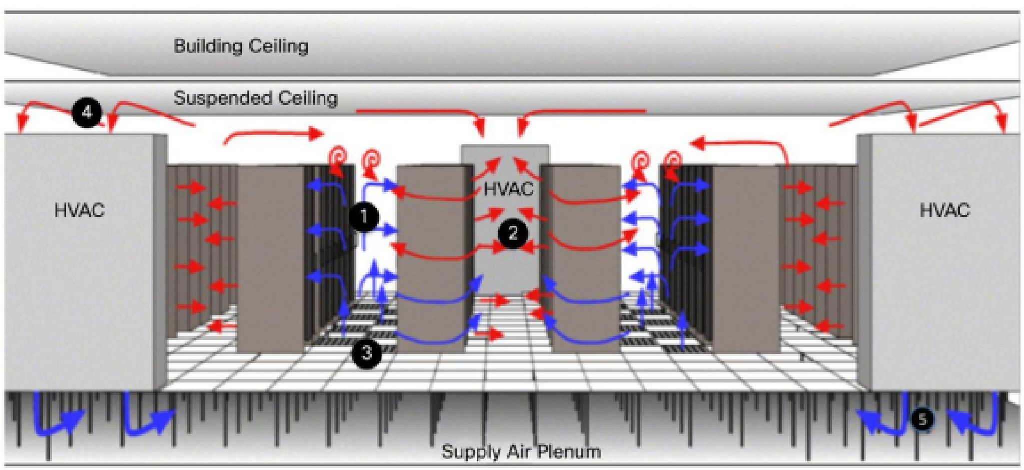


Figure 1-1 Typical thermal layout of a data center [4].

Figure 1-1, is a common data center architecture. A raised-floor is used to supply cold air and dropped-ceiling is used to return air with the CRAC. The airspace is divided into cold and hot aisle. A cooling system should include all levels of thermal management in a data center.

American Society for Heating Refrigeration and Air Conditioning Engineers (ASHRAE) TC 9.9 [5] Committee has developed a guideline for design, operation, maintenance, and efficient energy usage of data centers. The recommended temperature zones are from 18°C (64.4°F) to 27°C (80.6°F). The humidity should be less than 60%. Also, ASHRAE A3 envelope allows IT equipment to operate at 24˚C temperature and 85% relative humidity [3]. The operating ranges for both air and water cooling are illustrated in Figure 1-3.

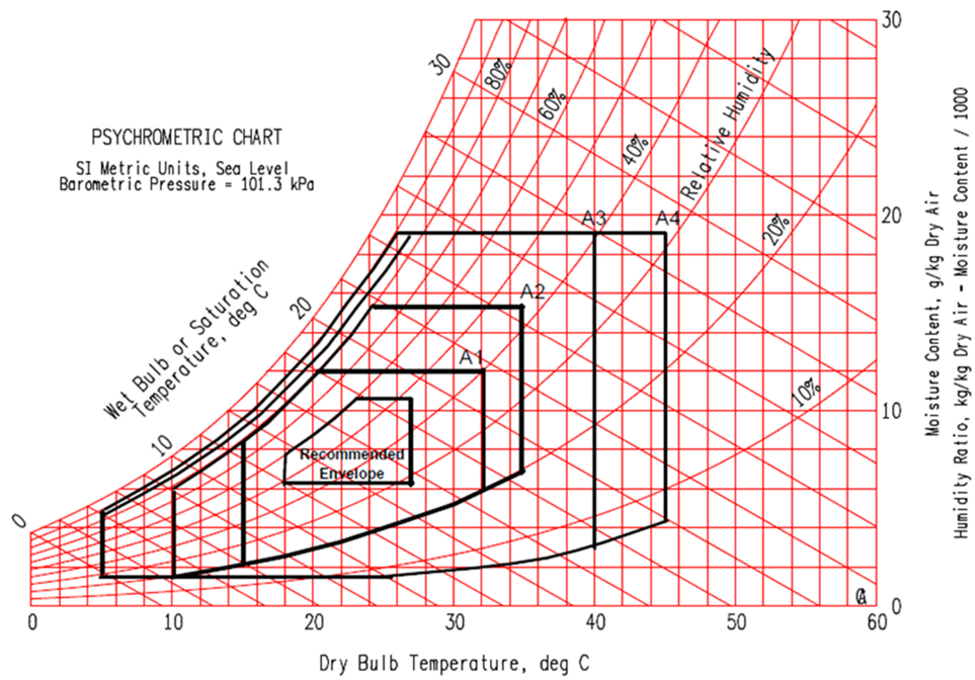


Figure 1-2 Environmental envelopes based on the class of data center [6].

The heat removal process in a server form chip level to ambient air includes design of effective heat sink, use of thermal interface materials and heat spreader, proper fan installation and ducting air-delivery pathway to the heat sinks. In a rack, cooling of servers depends on optimization of airflow rate comprising cabinets, walls and server conditions. Various arrangements are possible to improve the cooling efficiency by employing perforated tiles, hot & cold aisle containment, rear door heat exchangers etc. to deliver cold air from Computer room air-conditioning (CRAC) to servers. Some of the important factors for improving the efficiency are the positioning of fans, rear-door heat exchangers, racks and spacing between the servers (in a rack) and liquid cooling for high density servers etc. Various distribution configurations are used to control airflow delivery by air mixing and dispense air to the loads [7].

Liquid cooling has acceptance for cooling novel, high-powered microelectronic device [8]. To remove the heat more effectively liquid instead of air can be used. The cold plate is used to remove the heat dissipated by chips mounted on the glass ceramic substrate. As the module powers is increasing, liquid cooling is once again being considered for thermal management of microelectronic devices [9]. Utilizing heat carrying capacity, targeted delivery and lower transport power, liquid cooling can outcome the limitations of air cooling in non-uniform heat dissipation by the processors. Also, when operating at higher utilization, servers are more energy-efficient [10].

This study is devoted to understanding the impact of both reduced air and liquid flow rate on the thermal performance and energy consumption within a server which will be helpful in optimizing the hybrid cooling in a server.

1.5 Scope of Work

The objective of this work is to reduce cooling power by reducing flow rate of air and liquid in a hybrid cooled server and examine the temperature rise for deciding safe operation of servers. The overall objectives of this study are as follows:

- Evaluating performance of a hybrid cooled server during pump and fan failure.
- Understand the effects of high coolant inlet temperature on CPU die temperatures.
- Optimize the setup for efficient operation at raised water inlet temperatures.
- Reduce cooling power.

Chapter 2

Experiments

To observe the effect of raised inlet coolant and reduced air flow rate, the server was connected to DAQ, thermocouples and power meter. The experiments were performed to check for redundancy in total cooling power and evaluate the response and risks in the event of pump and fan failure scenarios.

2.1 Server under Study

The server for this experiment is Cisco UCS C220 M3 [11]. It's a 1U rack mount unit and the heat sinks are replaced with ASETEK cold plates for processors. The main components are listed below:

Main Components:

- Intel Motherboard
- CPU-Intel Xeon E5-2690 v2-3.00GHz (2, TDP of 135 W each)
- Memory-DIMMs (16, 96 GB)
- PCH
- HDD (8, 300GB each)
- Power Supply (650W)

Cooling:

- ASETEK Cold Plates (retrofitted on Processors)
- Server fans (40x40x56 mm, 5 fans)

The fans are positioned in front of DIMMs and CPUs on the inlet side of the servers. The fans in a typical server is controlled by a fan control logarithm which sets different PWMs based on die temperature of components such as CPUs, DIMMs and PCH.

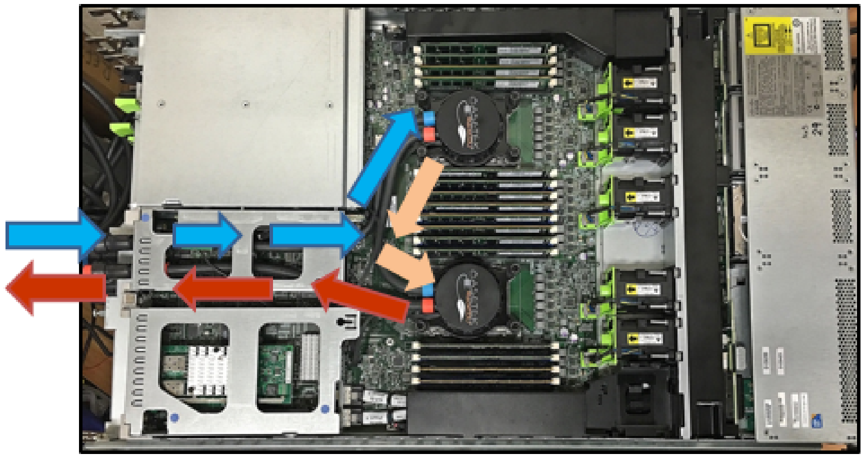


Figure 2-1 Cisco Server with retrofitted ASETEK Cold Plate

There are two pumps for circulating the coolant through a cooling loop and take away the heat from CPUs. Each pump is integrated with cold plates and consumes about of 4W each. The external loop for the liquid cooling of this setup consists of a liquid to air heat exchanger or radiator with axial two fans on top.

2.2 Test Setup

For the experiment, an isolated hybrid cooled server was used. The fans and pumps are powered from outside and the Pulse Width Modulation (PWM) signal generator controls fan speeds. The coolant from the server is cooled by a miniature dry cooler(radiator). The power supply supplies 12V to both radiator fans and server fans. Two penetrable K-type thermocouples are used to measure the inlet and outlet temperature of the coolant through Data Acquisition (DAQ) unit. The diameter of the thermocouple is negligible to the flow.

Figure 2-2 K-type Penetrable Thermocouple

The data from the thermocouples is collected. So, to perform the experiments with varied range of inlet coolant conditions, the inlet temperature for coolant was controlled by changing the PWM of radiator fans via LabVIEW and Arduino UNO. The Schematic for the test setup is shown in Figure 1.

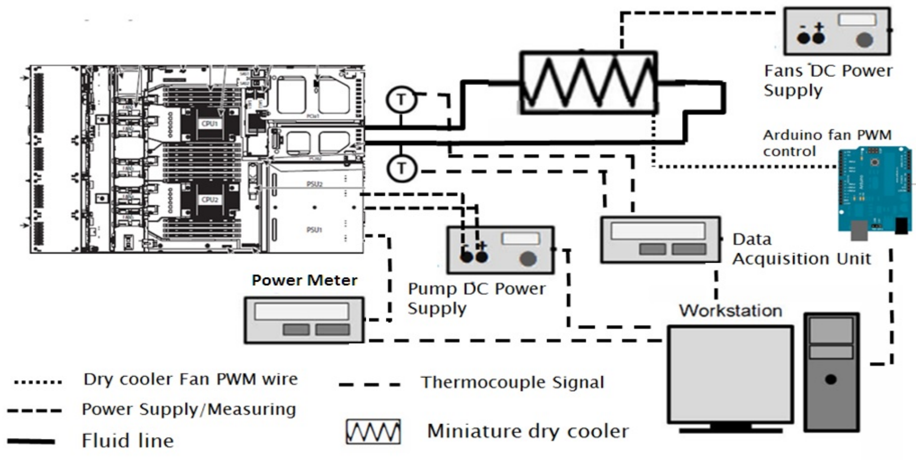


Figure 2-3 Schematic of test setup and data acquisition

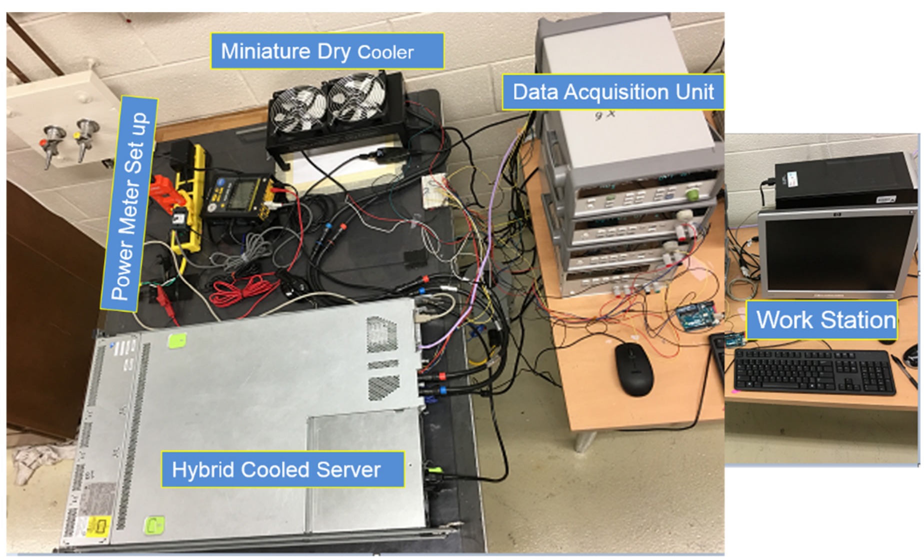


Figure 2-4 Experimental test setup and data acquisition

The LabVIEW code was used to control the inlet temperature of the liquid coolant. To control the temperature, the speed of the dry cooler fan was varied. First the inlet temperature was measured though a DAQ and the DAQ is connected to the LabVIEW code. According to that temperature LabVIEW sends a PWM signal to the fans through Arduino.

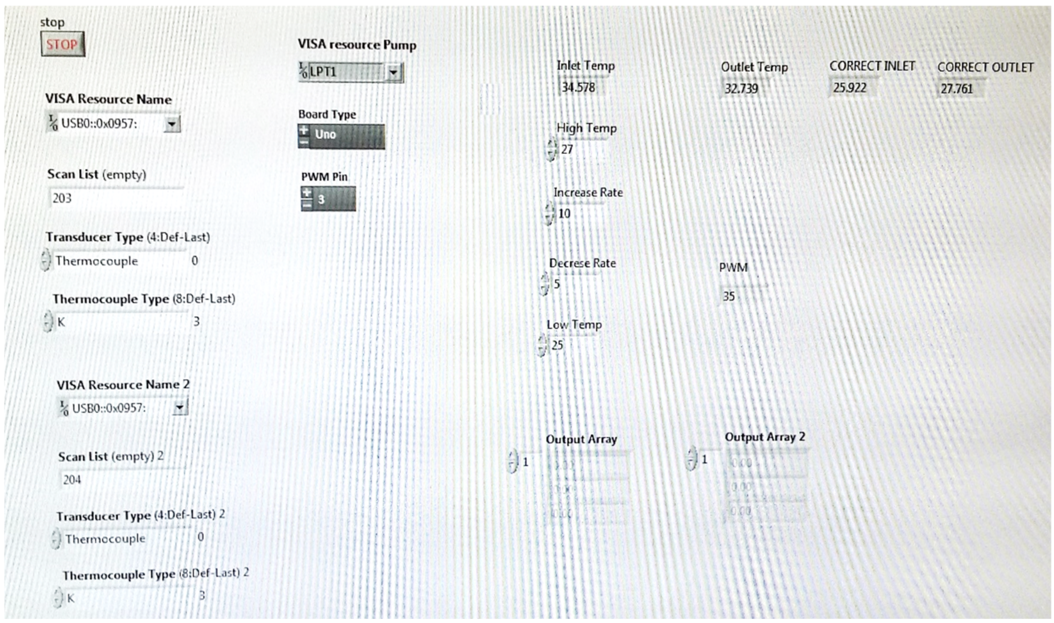


Figure 2-5 Front Panel of LabVIEW code

The power consumed by the server is collected by an external power meter (Yokogawa CW121) which is connected to the workstation. The power and RPMs for the pumps and fans are also measured by DAQ and Benchvue software. The power data is collected from time to time to monitor any unexpected change in power consumption of the server in response to the component temperatures.

2.3 Test Procedure

In an event of power failure, the most heat generating components needs cooling as UPS or generator is generally dedicated to providing power to IT equipment. The scope of the study is based on the effect of temperature rise of the server components to calculate the compensation in cooling in the event of failures. As the system is a hybrid cooled server, the main heat generating components are cooled by the inlet coolant and rest of the components such as DIMMs, MOSFETs, PCH and other auxiliary components are cooled by air drawn by the fans in the server. The server level study is performed in the sense to analyze temperature rise of individual components and find a baseline to predict safe service and operating condition for the failure scenarios on a server level. With respect to cooling, the failure scenarios were categorized as the failure of liquid cooling and air cooling.

The liquid cooling loop consist of two pumps connected in series and the coolant was cooled via an external heat exchanger (ASETEK miniature dry cooler). For failure scenarios, the cooling of CPUs is critical and pumps are driving force for the coolant in a loop. The pump failure scenario consisted of failing one of the pumps and observe the effect of temperature rise with different level server utilization. The inlet condition for coolant was also varied to characterize the system for high ambient conditions. The liquid cooling guideline for data centers provided by ASHRAE TC 9.9 [6] was followed for the experiments by varying inlet coolant temperature from 25°C to 45°C with an increment of 5°C. To evaluating the worst possible condition, the pump RPMs are varied as well for each coolant inlet temperature while running only one pump. Three RPMs are selected for the study and in each case, rpm is changed by setting a voltage (7V, 10V, 12V) in power supply unit. In each case, a computational workload of idle, 10%, 30%, 50%, 70%, 100% was applied for CPUs using a command line called “stress”, an internal tool for Ubuntu

[12]. The effect of pump failure on the temperatures of CPUs and other components are monitored and recorded by Intelligent Platform Management Interface (IPMI) tool [13] in UBUNTU operating system of the server. The sensor data are collected every 3 seconds and server power data was also collected to observe the changes in power consumption. High temperatures were achieved for maximum server utilizations and the tests are repeated three times to ensure repeatability and reduce errors. Temperature rise was studied to decide safe operating conditions for an individual server in the event of single pump failure and to understand the control parameters.

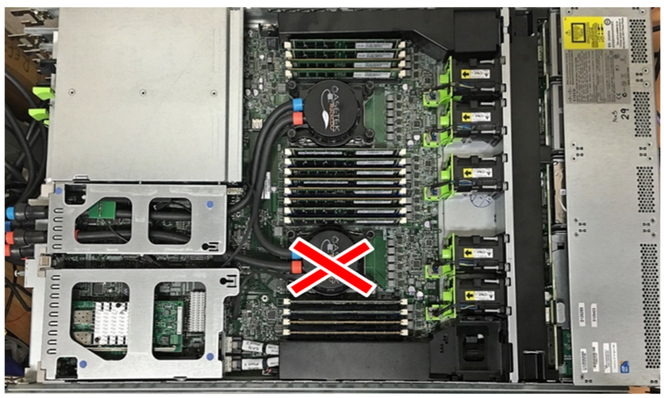


Figure 2-6 Pump failure scenario for downstream pump

There are five fans in the server and components other than CPUs are cooled by the airflow generated by the fans. Server fans are generally controlled by the built-in fan control algorithm which depends mainly on the high heat generating components. As this server is retrofitted with cold plates, the high generating component, CPUs was cooled by the liquid coolant. So, the main heat generating components for air cooling were PCH and DIMMs.

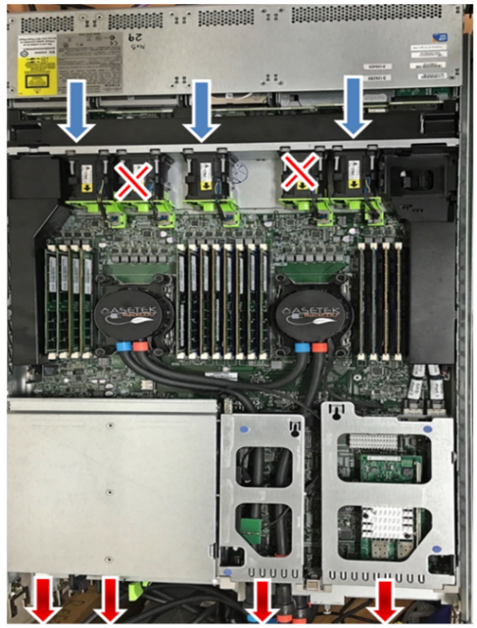


Figure 2-7 Fan Failure scenario for 2 fans.

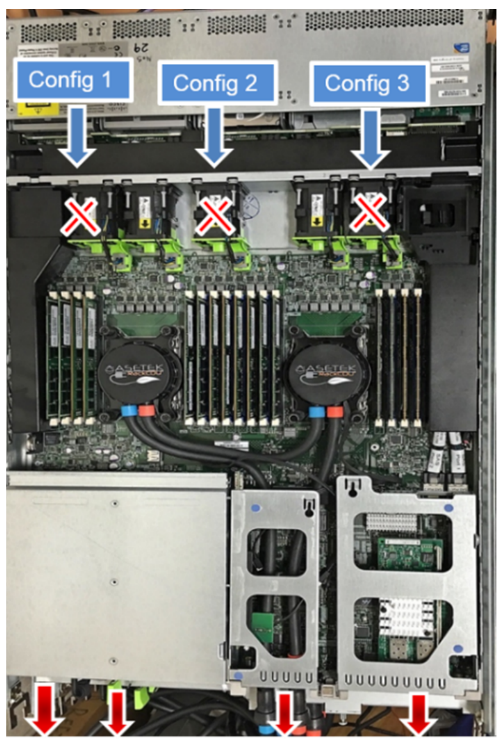


Figure 2-8 Fan Failure scenario for 3 fans

To understand the effect of temperature rise, fans are powered externally by a power supply and controlled by a PWM signal generating unit. For predictive fan failure scenarios, fans were disconnected one by one in each case. Fans in front of the pumps are turned off and later cases with fewer fans in front of the DIMMs are tried while keeping server utilization at maximum. In each case, the fan PWM was varied in five steps as 30%, 40%, 55%, 70%, 85% and 100% and maximum computational workload was applied for CPUs (100%) and memory is stressed up to 92% by a stressing tool "prime95" [14]. Full memory utilization was not applied as sensor data could not be collected for performance issues. The component temperatures are recorded by IPMI tool and each stress test continued for 15 mins when steady state conditions are achieved. The tests are performed twice to check for repeatability and errors.

Chapter 3

Results

3.1 Results for Pump Failure

The pumps are connected in series and provides coolant to the cold plates to cool the CPUs. Pumps are powered externally by a power supply. By changing the voltages, the pump rpm was varied and instantaneous rpms were measured through a DAQ. Three rpms were selected to run experiments ranging from 2500 to 3500 rpm. The test script for each case ran for 6 hours, with 30 minutes of stressing the CPUs at idle,10%, 30%, 50%, 70%, 100% and 30 minutes of idling in between.

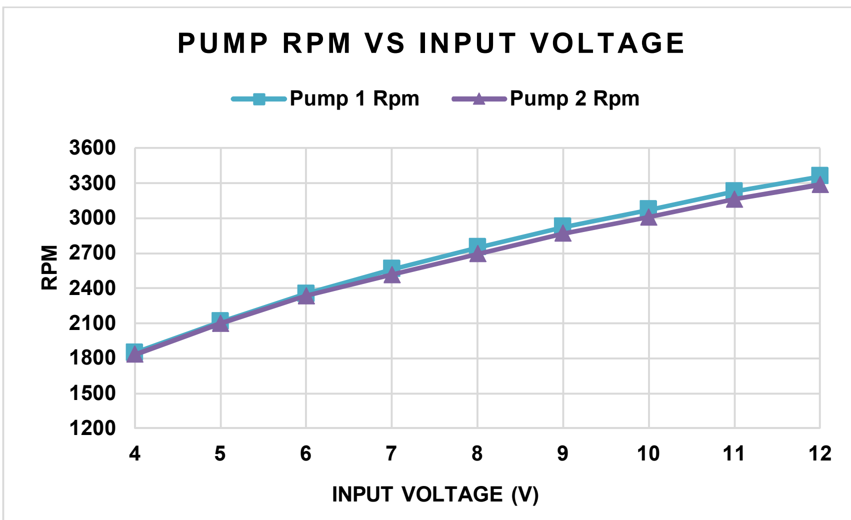
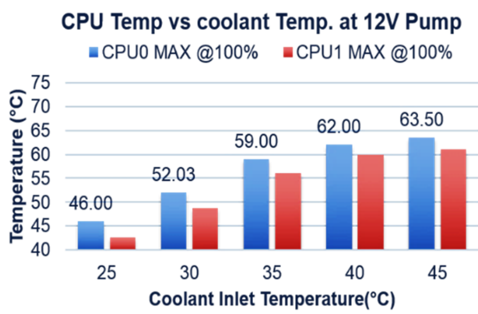
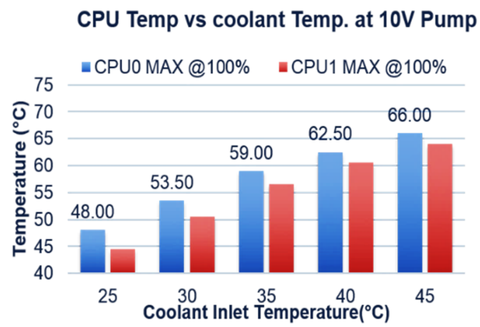


Figure 3-1 The variation in pump rpm for different voltages.

For initial testing, experiments are run for 2 pumps and the data was collected by IPMI tool. Each case was examined and CPU temperatures are plotted for different inlet temperature of the coolant to observe effect of warm water cooling with variable flow rate. The tests are repeated two times to ensure repeatability of data and changes in temperatures are recorded. The temperature for CPU0 (downstream) was found to higher than CPU1(upstream). The maximum temperatures were observed for maximum CPU utilization. As the CPU0 exhibited higher temperature, the worst-case scenario will be the situation when pump for CPU0 fails. So, the downstream pump was turned off and the experimental procedure was followed as before.





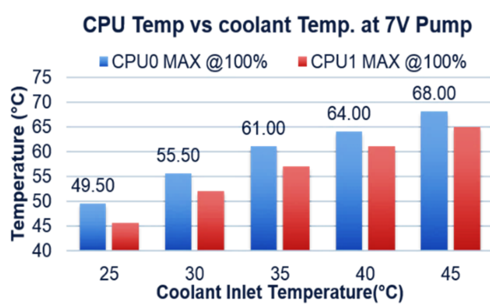


Figure 3-2 CPU Temp vs Coolant Temperature with 2 Pumps

By powering only one pump, same stress load has been applied and change in CPU core temperatures were monitored by “Psensor” tool [15]. Necessary precautions were followed while monitoring the CPU temperature. As the pumping power varies higher temperatures are observed for maximum CPU utilization. For failure of one pump in series, the maximum temperature for CPU0 was found to be 70.5°C at 45°C inlet and 7V supply voltage input (2500 rpm) which was for 7V power input (2500 rpm) and this case is the worst-case scenario. Due to pump failure, there is an average rise of 2.5°C in CPU temperatures. The temperature rise has been reducing from 25°C as we go to 45°C inlet coolant.

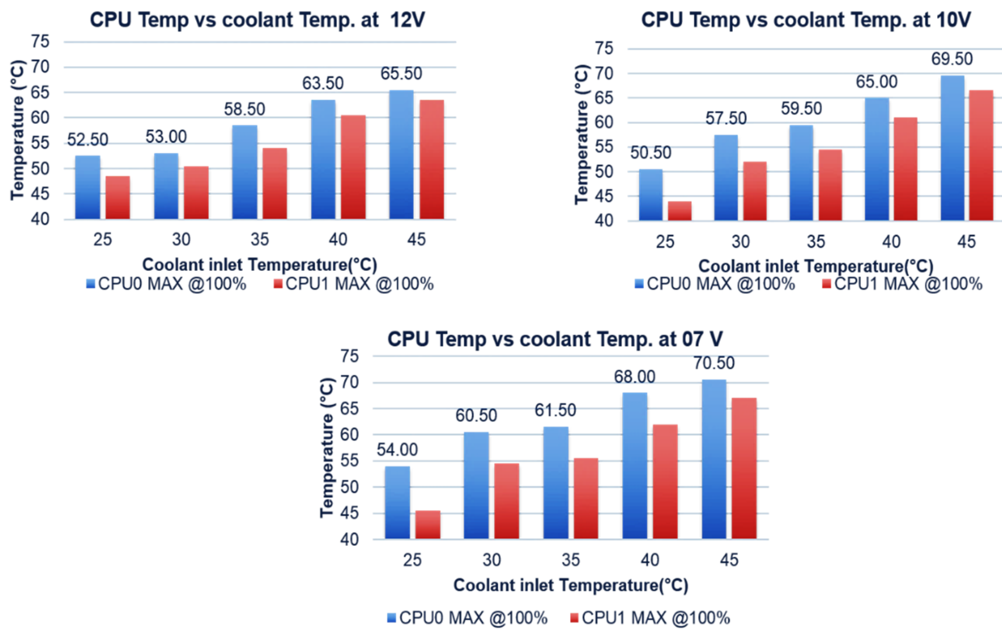


Figure 3-3 CPU Temp vs Coolant Temperature with Single Pump

3.2 Results for Fan Failure

The initial set of testing has shown that the fans are internally controlled based on the CPU temperatures. This setting is not applicable as the CPU are liquid cooled. Hence, for our analysis purposes, the fans were powered and controlled from outside. The study for fan failures is performed by turning off the fans and providing different PWM to the rest of the fans. The computational load was provided by “prime95” tool to both CPU and memory and sensor data for CPUs, DIMMs and PCH are collected by “IPMI” tool. The test was conducted for five (base line), four and three fans to predict the failure scenarios. In a previous study [16] it is found that a minimum flow rate of around 10 cfm is required for this server. Fans in front of the CPUs were turned off one by one for the case study of five and four fans. Each case was experimented for different fan PWM at 100%, 85%, 70%, 55%, 40%, 30%. Corresponding data is collected and plotted to observe the effect of reduced flow rate on the component temperatures as the fan number was decreased. Each test ran for 15 minutes until steady state is achieved for the temperatures and test are performed twice to ensure repetition of the data.

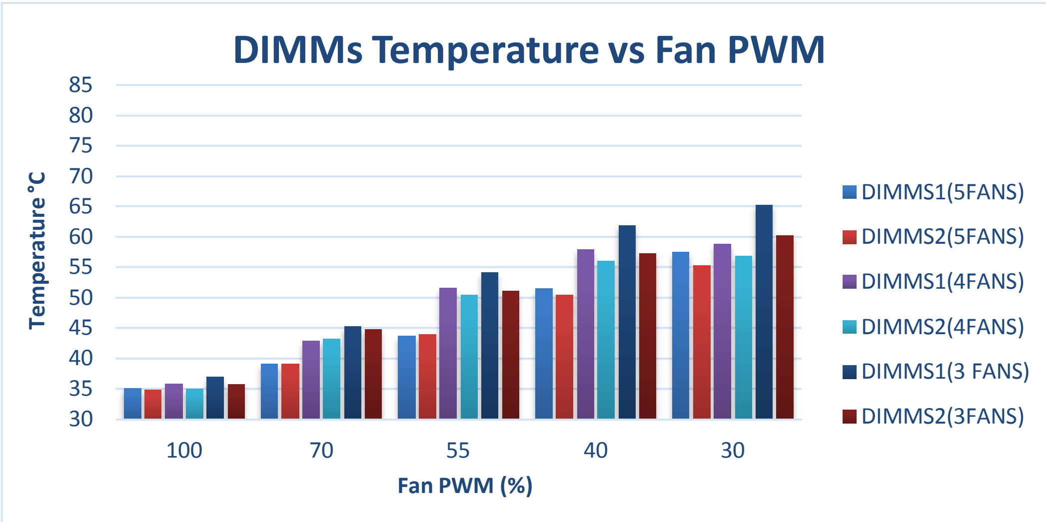


Figure 3-4 DIMMs temperature for 5,4,3 fan configurations at different fan PWM.

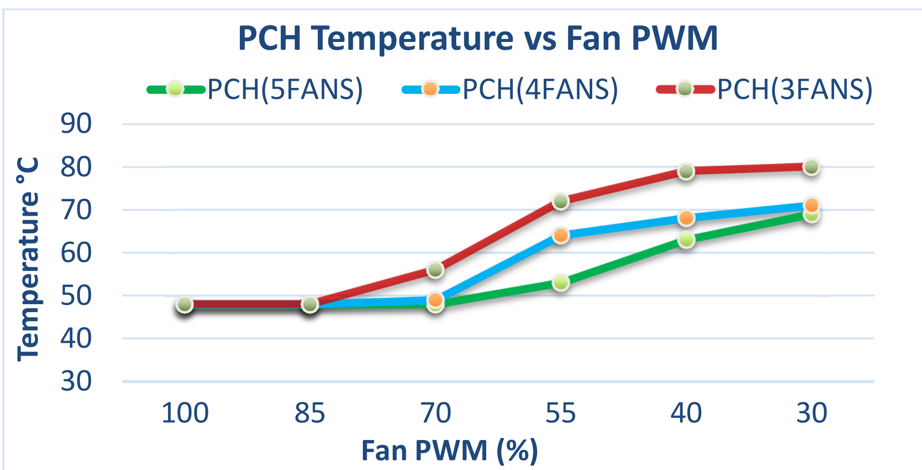


Figure 3-5 PCH Temperature for 5,4,3 fan configurations at different fan PWM.

Location of fans is another factor for fans in a 1U server and fans ahead of DIMMs are of prime concern for air cooling. So, considering the worst-case scenarios, the fans in line with DIMMs are turned off and fan PWMs are changed ranging from 30% to 100%.

Three cases (config 1,2,3) are studied by turning off fans for each side and in the middle, one at a time.

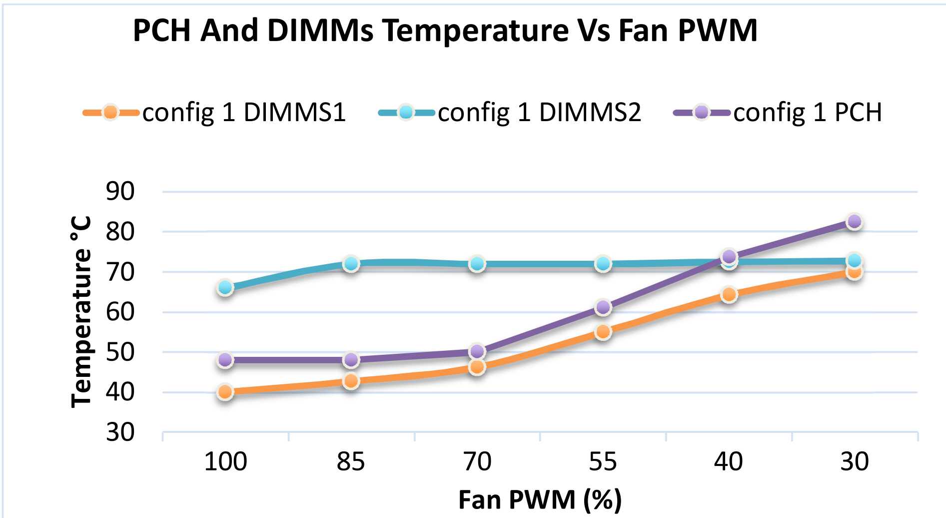


Figure 3-6 Fan failure scenario with config 1

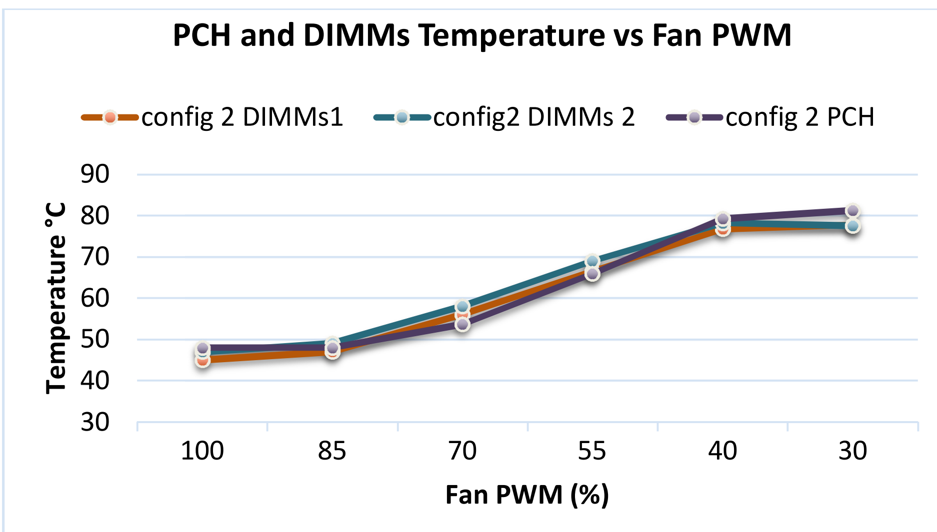


Figure 3-7 Fan failure scenario with config 2

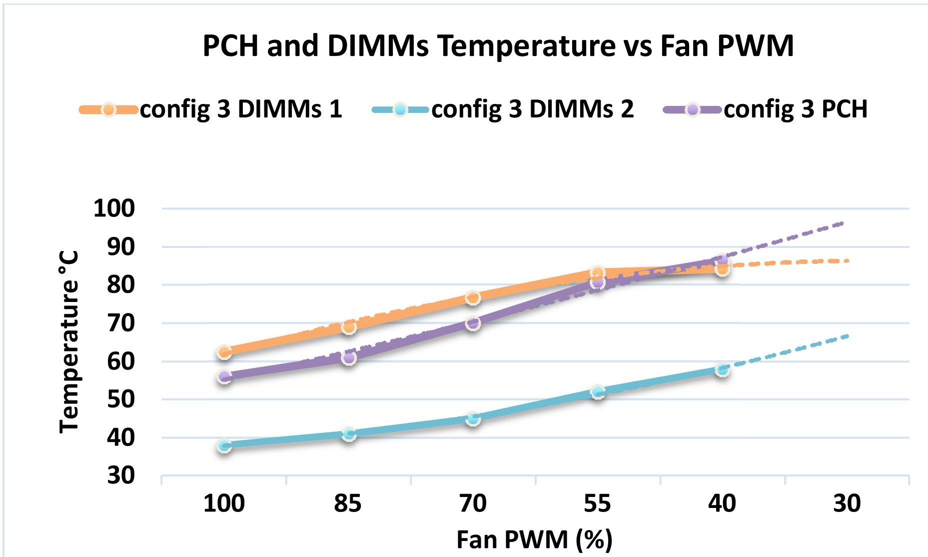


Figure 3-8 Fan failure scenario with config 3

Each test continued for 15mins until steady state was achieved. For third case (config 3), 30% PWM may cause damage to PCH as reduced air flow has caused a rapid rise in temperature close to the critical temperature at 40% PWM. The data for the lowest PWM has been extrapolated. Sensor data was collected with IPMI tool. Maximum DIMMs temperature is calculated from the averaged values over the last five minutes of the test run.

Chapter 4

Transient Analysis

4.1 Full Pump Failure Scenario

The Pumps in this hybrid cooling generates flow rate for the coolant to pass through the cold plates and carry away the heat from both the CPUs. The CPU temperature depends mainly on utilization and rate of heat removal. The flow rate in conjunction with cold plate heat exchanger can remove heat from the surface of the processor very effectively [17]. Turning off one of the pumps showed the rise in temperatures for coolant as well as CPU junction temperature. But in the event of total cooling power failure, pumps will not be able to provide the amount cooling needed and the CPUs may be damaged due to overheating. In this experiment, the server was tested to approach the maximum CPU die temperature and calculate the time for temperature rise and monitor the pattern

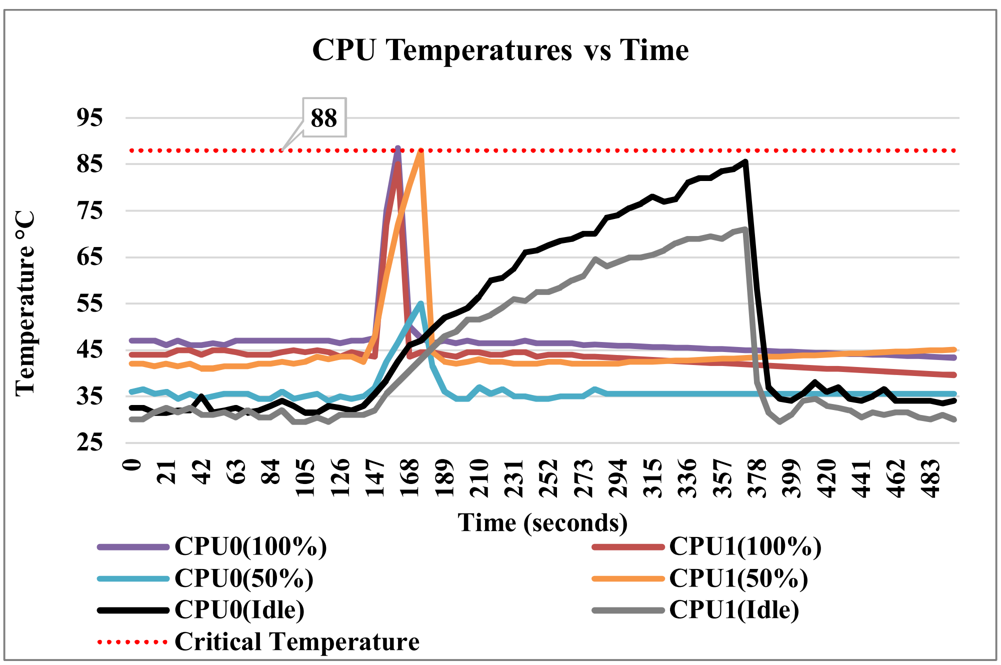


Figure 4-1 Full pump failure scenario with CPU temperatures vs time

Both the pumps are turned off and instantaneous sensor data was collected. The CPU was stress from idle to max with 25% increase in each step and results are plotted for idle, 50%, 100% CPU utilizations as shown in the graph. As the utilization increased, the die and junction increased and the lines for temperature rise for separate utilization can be noticed from the graph (right to left). Taking critical temperature into consideration, power supply to the pumps were controlled by monitoring the CPU core temperatures.

4.2 Full Fan Failure Scenario

Servers are provisioned for the failure scenarios of fans in air cooling and safety features for saving the components from such critical conditions can be found in reference [18]. But considering the worst case when all the fans fail, the temperature rise is expected but the risks are far less for hybrid cooled servers compared to air-cooled servers. To examine the condition experimentally and measure power consumption during the event, all the fans were powered off externally and the maximum computational load was given with “prime95” for both the CPU and memory. As the CPU was being cooled by the liquid coolant, the high temperature was not noticed for CPUs for liquid cooling. The fans were stopped around 147 seconds and powered again around 780 seconds. Temperature of other air-cooled components like PCH (blue) and DIMMs (labeled as in reference [11]) were plotted for transient analysis as shown in figure.

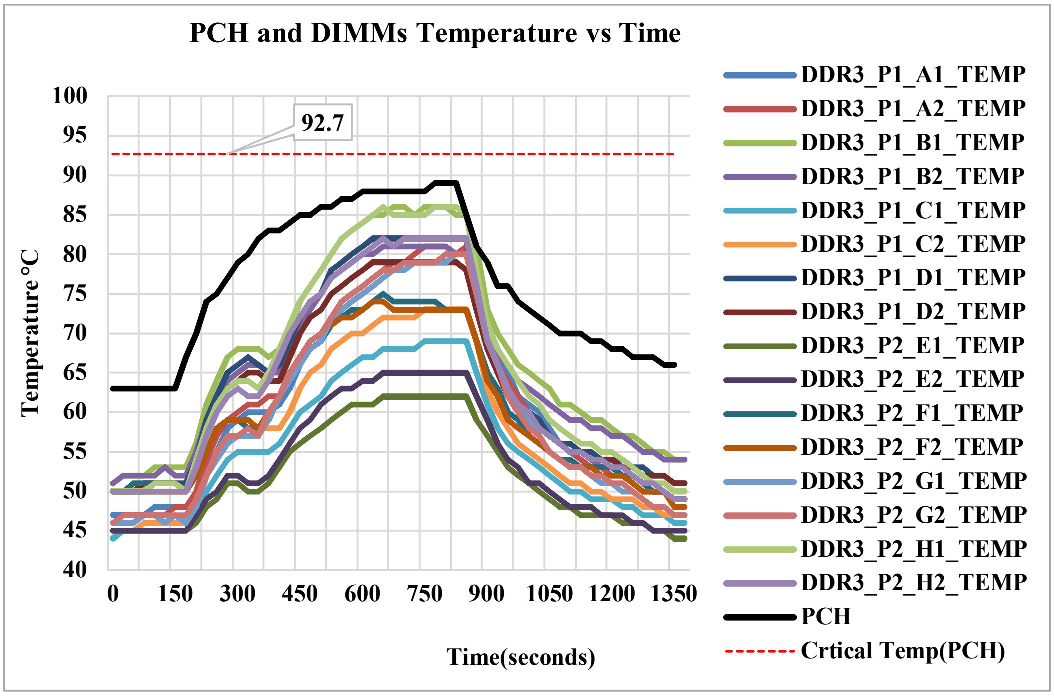


Figure 4-2 Full fan failure scenario with DIMMs and PCH temperatures vs time.

Chapter 5

Power Consumption

5.1 Power consumption in Pump Failure Scenario

Typical data centers are defined by the floor area of the IT facility and the power that the facility can support per unit area [19]. Calculating the power consumption from each of the server at various conditions provides a guideline for calculating the Power Usage Effectiveness (PUE) and characterize the data center cooling efficiency. In hybrid cooling most of the produced IT heat is transferred in liquid cooling loop, amount of air cooling needed for other components is reduced. But quantifying the amounts is beneficial for calculating total cooling power. As the experiment is conducted with an isolated server, the pPUE guideline [20] is followed to calculate total cooling power required.

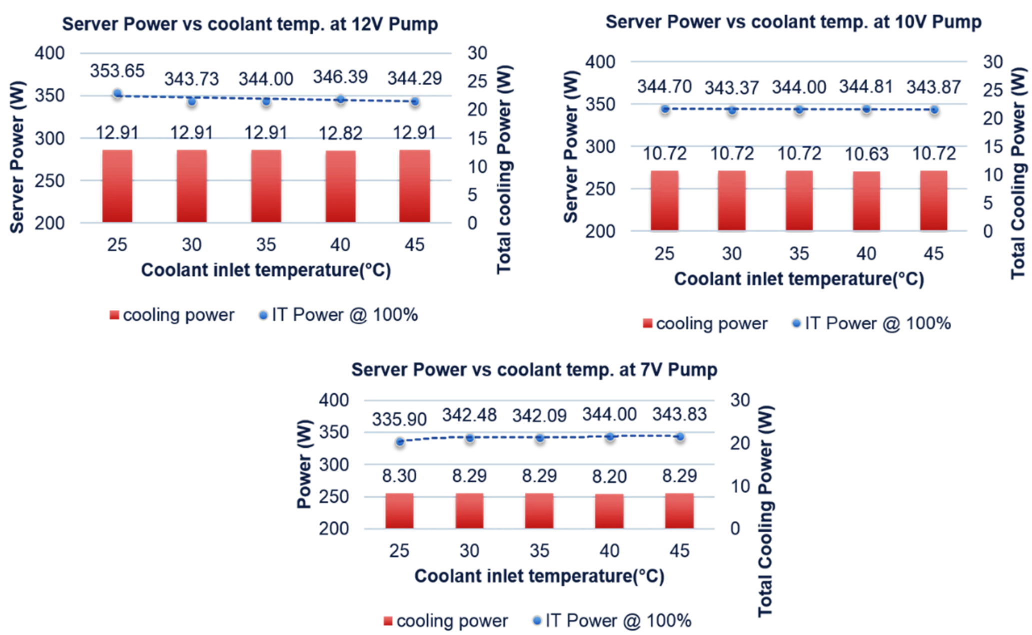


Figure 5-1 Power consumption at Maximum CPU Utilization with 2 pumps

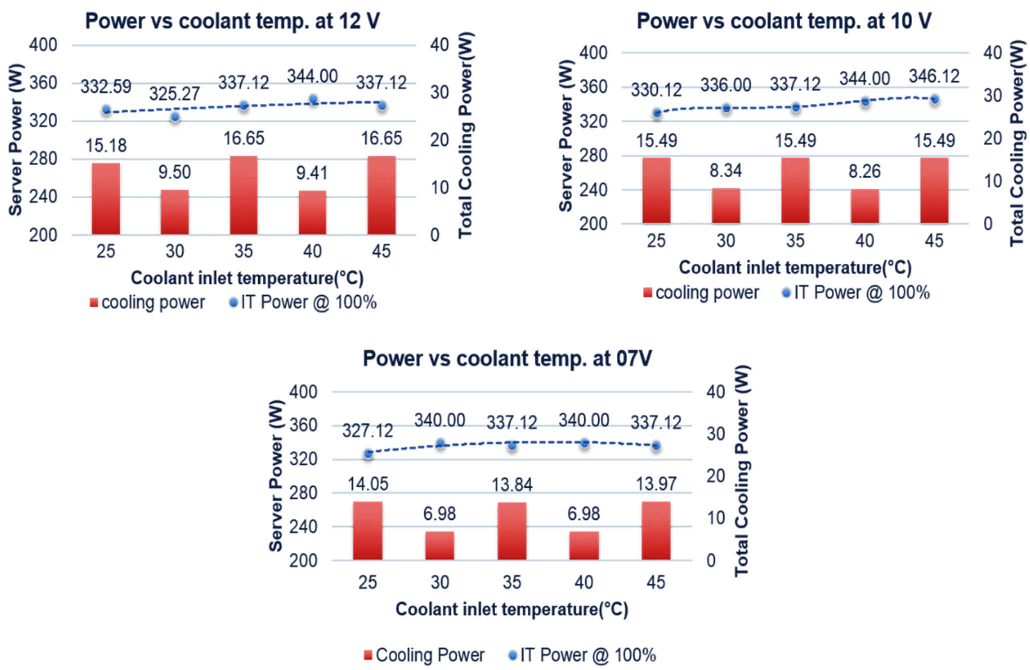


Figure 5-2 Power consumption at Maximum CPU Utilization with 1 pumps

The total cooling power includes cooling power (fan power and pumping power) required within the IT equipment boundary. The fan and pumping power was collected by Benchvue software via Agilent Power Supply unit and added together to calculate the total cooling power. The powers consumption shows higher values when both CPUs and memory are utilized to the maximum. The graphs for pump and fan failure shows different but almost uniform power consumption for different cases. The fans were controlled in two inlet conditions (30 °C and 40 °C) for single pump while for other inlet conditions, the fans were controlled internally by the fan control algorithm provided within the server.

5.2 Power consumption in Fan Failure Scenario

The fans were controlled externally and gradual change in PWM from 30% to 100% was provided to observe the rise in component temperature. At the same time power consumption was recorded as before.

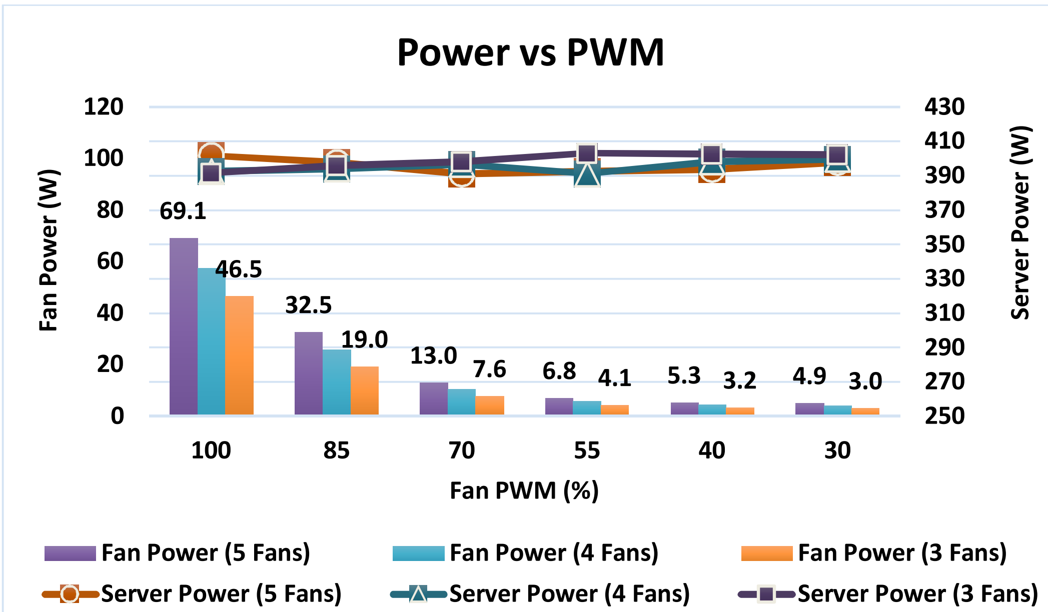


Figure 5-3 Power Consumption in Fan Failure Study

The failure study for fans provided the temperatures and the reduction in fans caused reduction in total cooling power. The previous study also showed choosing three over five fans is advantageous and in failure scenarios of any of three fans the server fans must be replaced to help the server operate smoothly. Here, the graph is plotted for five, four and three fan setup and respective server and cooling power is represented for different PWM. The server power consumption was around 400W with 3-4% variation. The fan power consumption decreases as fan PWMs decreases. A reduction of **55.88%** in fan power observed comparing 5fans (55% PWM) with 3 fans (30% PWM).

5.3 Reduction in Pumping Power

The pumping power was calculated from the power supply unit. Two pumps require approximately 7.17W and single pump requires about 3.76 W. Again, changing the rpms of pumps from 2500 rpm to 3300rpm and comparing single pump (7V) to two pumps (12V), maximum 81.32% reduction in pumping power is observed. With the use of only three fans and single pump, total cooling power reduces from 18.37W to 9.42W with a reduction of 45.93%. The reduction in total cooling power with reduction in pumping power is shown here for clear understanding.

Table 5-1 Reduction in Pumping Power

Voltage (V)	**Two Pumps Power (W)**	**Single pump Power (W)**	**Total cooling power (two pumps)**	**Reduction (two Pumps)**	**Total cooling power (Single pumps)**	**Reduction (Single pump)**
12	7.18	3.77	12.91	--------	9.5	26.41%
10	4.99	2.61	10.72	16.96%	8.34	35.39%
07	2.56	1.34	8.24	**36.17%**	6.98	**45.93%**

Chapter 6

Conclusion and Future work

6.1 Conclusions on Failure Scenarios

In this study, different possible failure scenarios for an isolated hybrid cooled server has experimented and warm water cooling is employed to experimentally investigate the worst-case scenarios and opt for system redundancy. The system uses both liquid and air cooling and optimizing the cooling can provide a way to reduce cooling cost. Comparative analysis of two pumps to one pump shows the rise in CPU temperatures but the temperatures remain within safety zone well below the critical temperature even in the worst possible condition with warm water cooling. The transient analysis with pump failure shows less time than that for fan failure scenario. In other words, maintaining the pumps is more crucial for this hybrid cooled server. The time for response, repair and replacement can be calculated from transient analysis and safety measurements for hybrid cooled servers. Here, reducing the cooling power is another aspect to improve cooling efficiency. The power consumption analysis describes a gap of watts which shows a possible reduction in cooling power (45.93%) is possible on a server level without causing overheating and hampering IT equipment life. Pumps with low flow rate can be used for this cooling loop. But a rack level study can help to validate the approximation for a typical data center. So, future study can be based on the pressure drops and flow rate required for optimum cooling to evaluate the parametric performance of the servers running over a wide range of operational conditions.

6.2 Future Work

- Develop a valid prototype of **flow network model.**
- Determination of **optimum flow rate** and **pressure drop** (max & min).
- Perform a rack level study to validate the approximations.
- A comparative study with air cooled servers.

References

[1] J. Koomey, "My new study of data center electricity use in 2010," 31 July 2011. [Online]. Available: http://www.koomey.com/post/8323374335.

[2] Prabjit Singh, Levente Klein, Dereje Agonafer, Kanan Pujara;, "Effect of Relative Humidity, Temperature and Gaseous and Particulate Contaminations on Information Technology Equipment Reliability," in *InterPACK/ICNMM*, San Francisco, 2015.

[3] *Thermal Guidelines for Data Processing Environments– Expanded Data Center Classes and Usage Guidance Whitepaper prepared by ASHRAE Technical Committee (TC) 9.9 Mission Critical Facilities, Technology Spaces, and Electronic Equipment,* 2011.

[4] Cisco White Paper, "Cisco Unified Computing System Site Planning Guide: Data Center," 2017. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/data-center-virtualization/unified-computing/white_paper_c11-680202.pdf. [Accessed 19 5 2017].

[5] "2008 ASHRAE Environmental Guidelines for Datacom Equipment," *ASHRAE TC 9.9 committee,* 2008.

[6] ASHRAE TC 9.9, Thermal Guidelines for Data Processing Environments, Atlanta: American Society of Heating, Refrigeration and Air-Conditioning Engineers Inc., 2011.

[7] K. D. a. N. Rasmussen, "Choosing Between Room , Row , and Rack-based," p. 18, 2012.

[8] B. Agostini, M. Fabbri, J. E. Park, L. Wojtan, J. R. Thome and B. Michel, "State of the art of high heat flux cooling technologies," *Heat Transfer Engineering,* vol. 28, no. 4, pp. 258-281, 2007.

[9] R. R. Schmidt, "Liquid Cooling is Back," 1 August 2005. [Online]. Available: http://www.electronics-cooling.com/2005/08/liquid-cooling-is-back/. [Accessed 23 January 2015].

[10] J. Fernandes, S. Ghalambor, D. Agonafer, V. Kamath and R. Schmidt, "Mutli-Design Variable Optimization for a Fixed Pumping Power of a Water-Cooled Cold Plate for High Power Electronics Applications," in *IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems*, San Diego, CA, USA, 2012.

[11] Cisco, "Cisco UCS C220 M3 High-Density Rack Server (Small Form Factor Disk Drive Model)," Cisco, 16 05 2017. [Online]. Available: http://www.cisco.com/c/dam/en/us/products/collateral/servers-unified-

computing/ucs-c-series-rack-servers/c220m3-sff-specsheet.pdf. [Accessed 19 05 2017].

[12] UBUNTU, "Linux man Page, "Ubuntu Manuals”," 2010. [Online]. Available: http://manpages.ubuntu.com/manpages/zesty/man1/stress.1.html. [Accessed 19 05 2017].

[13] UBUNTU, "ipmitool - utility for controlling IPMI-enabled devices," UBUNTU, 2010. [Online]. Available: http://manpages.ubuntu.com/manpages/zesty/man1/ipmitool.1.html. [Accessed 19 05 2017].

[14] Mersenne Research, Inc., "Prime95 Version 28.9 - CPU stress tool," [Online]. Available: https://www.mersenne.org/download. [Accessed 19 05 2017].

[15] "Psensor, A GRAPHICAL TEMPERATURE MONITORING APPLICATION," [Online]. Available: www.wpitchoune.net/psensor. [Accessed 19 05 2017].

[16] M. M. Islam, *xperimental Characterization of Hybrid Cooled Cisco Servers Including the Effect of Warm Water,* University of Texas at Arlington, May 2016.

[17] Asetek, "Asetek To Offer Liquid Cooling Solutions For Increasing Performance, Density And Energy Efficiency In Data Centers," 28 December 2011. [Online]. Available: http://www.asetek.com/press-room/news/2011/asetek-to-offer-liquid-cooling-solutions-for-increasing-performance-density-and-energy-efficiency-in-data-centers/. [Accessed 19 05 2017].

[18] Cisco, "Cisco UCS C220 Server Installation and Service Guide," [Online]. Available: www.cisco.com/c/en/us/td/docs/unified_computing/ucs/c/hw/C220/install/C220/replace.html. [Accessed 19 05 2017].

[19] R. Z. R. L. Brandon A Rubenstein, "Hybrid cooled data center using above ambient liquid cooling," in *12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, 2010.

[20] D. A. A. F. Victor Avelar, "PUE™: A Comprehensive Examination of The Metric," [Online]. Available: www.datacenters.lbl.gov/sites/all/files/WP49-PUE%20A%20Comprehensive%20Examination%20of%20the%20Metric_v6.pdf. [Accessed 19 05 2017].

Biographical Information

Uschas Chowdhury, has received his Bachelor of Engineering degree in Mechanical Engineering from the Bangladesh University of Engineering and Technology (BUET), Dhaka, Bangladesh. He has completed his Masters in Mechanical Engineering from the University of Texas at Arlington, Texas, USA in May 2017.

He worked as a Teaching Assistant for an undergraduate course in his second Semester and during his Master's, he also worked on projects under Dr Dereje Agonafer in The University of Texas at Arlington. He has associated himself with various industry collaborated research projects and studied various topics like air cooling, liquid cooling and direct/indirect evaporative cooling etc. He has gained theoretical background knowledge of CFD and worked data center specific CFD codes like 6SigmaET, 6SigmaRoom, ICEPAK and FloTHERM. He has worked in experimental and CFD characterization of Cisco servers optimization in 6SigmaET by variation in fan positions, thermal power optimization technique and determination of maximum die temperature.