

The Effect of Data Center Temperature on Energy Efficiency

Michael K Patterson
Intel Corporation
2111 NE 25th Avenue
Hillsboro, Oregon, USA, 97124
Phone: (503) 712-3991
Email: michael.k.patterson@intel.com

ABSTRACT

Server's capabilities are increasing at or beyond the rate of performance improvement gains predicted by Moore's Law for the silicon itself. The challenge for the Information Technology (IT) owner is housing and operating all of this computational power in the Data Center. With more computational power in each unit volume, the industry is experiencing a significant increase in power density and hence a greater cooling challenge. The ability to now tackle computational tasks that were previously unattainable has driven energy costs to new levels. Methods to reduce the energy used in cooling these machines are being studied throughout the industry. One of the areas being considered is increasing the Data Center server ambient inlet temperature. ASHRAE [1] suggests a recommended limit of 20 – 25°C for the most advanced Data Centers. There is a belief that operating at the high end of this range or above it will reduce the total power use in the Data Center by making the cooling system more efficient. A thermodynamic analysis clearly indicates that increasing the temperature of the high temperature heat source, while holding the lower temperature heat sink constant, will give an efficiency gain to the heat removal from the system. Unfortunately the simple model does not capture all of the components of the overall system and may lead to an erroneous conclusion. In fact, increasing the ambient temperature can lead to an increase in power usage of some components and systems in the Data Center as temperature goes up. The overall room energy use may only go down marginally or may even go up at warmer temperatures. In this paper we examine the complete energy picture from the utility connection to the rejection of heat from the facility to the outdoor environment and look at the impact an increased ambient temperature will have on each component in that chain. This analysis indicates that there is an optimum temperature for Data Center operation that will depend on each Data Center's individual characteristics, include IT equipment, cooling system architecture, Data Center location (e.g. outside ambient conditions), as well as other factors. Additional impacts of an increasing ambient inlet temperature, such as reliability issues and operational complexity are also discussed. It is concluded that simply raising the ambient temperature in the Data Center may not have the desired effect of energy use reduction.

Keywords

Server, Thermal Management, Fan Power, Economizer, PUE, Energy Efficiency, Room Temperature

NOMENCLATURE

A	area (m ²)
d	diameter (m)
h	Convective heat transfer coefficient (W/m ² K)
k	Conductive heat transfer coefficient (W/mK)
L	length scale (m)
Nu	Nusselt number
PUE	Power Usage Effectiveness
q	Heat transfer (watts)
Re	Reynolds Number
Pr	Prandtl Number
T	Temperature (C)
V	velocity (m/s)

Greek symbols

μ	viscosity (kg/sm)
ρ	density (kg/m ³)

Subscripts

air	associated with the ambient air
amb	ambient
eff	effective
heatsink	associated with the heat sink
junction	silicon peak temperature
outdoor	outside heat sink

Superscripts

n, m	equation variable coefficients
------	--------------------------------

INTRODUCTION

The cost to buy a unit of computational capacity continues to drop; driven by Moore's law at the silicon level and the ability of the server manufacturers to build more and more capable platforms. While end-users now can afford to fill a data center with servers they are often challenged to pay for the operations of that same data center. The cost to operate the server and associated cooling systems over the server's life is now close to parity with the cost to purchase the server. The costs associated with the increasingly large installed base are driving the end-user to find ways to reduce the operational cost wherever possible. The cooling of these systems is one area of exploration. Modern Data Center's could have a Power Usage Effectiveness (PUE) of 1.6 to 2.2. [2], [3]. The definition of PUE is

$$PUE = \frac{\text{Total Power}}{\text{IT Power}}$$

Where total power is comprised of IT power, power delivery losses, lighting, and cooling. The percentage of cooling varies but can be up to 50% [2] (or even higher in poorly designed or operated data centers). With cooling a significant operational cost, efficiency gains in that system are being explored and evaluated. In this paper we look at the impact of data center temperature on the overall efficiency of the total IT and utility infrastructure system. ASHRAE [1] recommends that the most critical data centers be maintained between 20 and 25°C, with an allowable range of 15 to 32°C. This temperature represents the temperature at the inlet to the IT equipment. Certainly temperatures in the space may be considerably hotter than this, particularly on the discharge side of the servers. But as long as the inlet side is maintained in the right range, the IT equipment will operate properly. All mainstream IT equipment is designed to operate with in this band. New Data Centers also are typically designed to operate and control room supply air temperatures. The choice of the room setpoint can be a key factor in cooling system's energy use and room operation.

Occasionally data centers are designed to operate with temperatures as low as 15 or 16°C based on a principal that colder is better; for compute performance; reliability, and stability. In addition, some data centers are designed to run at the warmer temperatures above 25°C up to 32°C with the goal of saving energy. However the majority are still within the recommended range (vs. the wider acceptable range). The scope of this paper is a consideration of the energy efficiency of the data center as a function of temperature inside that acceptable range. Any data centers outside that range (15-32°C) are either specifically designed for a unique situation or the operations need to be corrected prior to any optimization for energy efficiency.

The economic incentive to improve cooling efficiency has led to recommendations regarding the correct set point or changing of that setpoint that do not necessarily hold up. We consider the impact of temperature on the entire power use stack-up to demonstrate the total impact at the utility meter for the end-user. To this end we consider the power delivery including the uninterrupted power supply (UPS), and power distribution. We also evaluate the impact of temperature on the server itself and the power use in the server power supply unit (PSU), voltage regulators, central processing unit, memory, spinning media, and internal fans. Finally the power used by the cooling system is considered, from the computer room air-handling unit (CRAH), chilled water system, and the chiller/cooling tower subsystems. All of these areas add to the electrical energy used by the data center and as such the goal of the end-user should be to minimize total energy rather than optimizing any individual component or subsystem. Specific, limited optimizations such as this may actually have an adverse efficiency effect on the remainder of the system. Looking at the total energy use ensemble allows us to suggest where and when it makes sense to modulate room temperature to gain energy use improvements.

While a detailed analysis is outside the scope of this work, the temperature of the Data Center will also have impacts on the operations and reliability of the Data Center and supported equipment.

PROBLEM DESCRIPTION

The energy impact of modifying the room ambient set point with-in the ASHRAE allowable limits (15-32°C) is considered. All aspects of the energy use by the Data Center, the IT equipment, and the utility infrastructure must be comprehended to determine the total impact on energy consumption. We consider multiple components in the analysis to determine the relationship of

$$Power \propto f(T)$$

Each component in the overall energy use chain must be analyzed for its reaction to a change in temperature as well as any potential interrelationships that may exist.

There are numerous data center designs and infrastructure implementations. Each DC will react somewhat differently to changes in temperature, so it is the intent of this paper not to define a single relationship but rather to demonstrate a methodology and the breadth of analysis necessary to evaluate potential changes in room temperature. The PUE metric will be shown to be useful in the analysis of the energy use in the data center, with the value of the PUE impacting the potential savings.

Motivation

Many suggestions have been made in industry recently about raising the room temperature as a quick and simple method to reduce energy consumption. This change and the potential savings in the energy used in cooling may actually drive higher energy use in IT equipment and may not be a good trade off. The primary motivation of this work is to provide a detailed look at the overall effect on energy of these types of temperature changes.

Misconceptions

Energy use in all types of buildings is a concern and an area of focus. A California Energy Commission (CEC) guide [4] for savings for businesses suggests that a savings of "as much as 2 percent of your air conditioning costs for each one degree that you raise the thermostat." This is a basic well understood phenomenon. Note that the savings are only even mentioned in the air-conditioning costs, and not the full cooling cost. The other end of the spectrum of claims can be seen as well. Consider a recent statement in an on-line IT trade journal [5] in an article titled "Going Green in the Data Center". The article claimed savings by running buildings warmer of 5% of the total energy cost for each 1 deg F the set point is increased. From this, one might infer that changing the set point in the Data Center from 20°C to just 23°C could net the building owner a 25% total energy reduction for the Data Center. Clearly if this

were the case everyone would have done it by now.

The magnitude of the rule of thumb energy savings for increasing ambient temperatures has two primary components. The first is a reduction in cooling load. One of the primary loads in a commercial building is heat gain from outside through heat conduction thru the walls and building envelope, solar gain, and infiltration. The external heat gain can be on the order of magnitude of 50% of the total load [6]. The cooling system must then remove the heat from the building. The energy it takes to do this is proportional to the differential temperature between the outside environmental temperature and the indoor ambient. As the ambient temperature approaches the outdoor temperature, the external cooling load thru the building envelope decreases significantly. This is a large part of the savings obtained with a warmer room temperature. The second portion of the efficiency gain comes from the cooling system itself. Increasing the room ambient allows the source temperature to be higher, increasing the efficiency of the cooling system.

In the majority of commercial buildings raising the room set point temperature both a) increases the efficiency of the cooling system and b) reduces the load.

It is informative to compare the nature of the cooling loads in a Data Center vs. a typical commercial building. An EPA report on energy consumption across the United States [6] showed that the typical energy consumption of a commercial building was on the order of 2 watts / sq ft. Current Data Centers [7] are being designed and built at much higher densities, typically 100-500 watts / sq ft. While in a typical commercial building the heat gain from the external was shown to be order of magnitude of 50%, in a typical data center the external load is a very small fraction and as such, the "rule-of-thumb" savings potentials need to be seriously questioned. While a) the cooling plant efficiency gain is still there b) the reduced load is not. The Data Center load is nearly all internal heat gain from electrical energy dissipation and does not fit the model the savings potential was built upon.

The energy savings for modifying ambient temperature must be developed from a detailed analysis of the data center and each component's response to a change in temperature.

Analysis Method

We consider the individual components and then we consider a higher level view and review the overall energy impact, as well as how PUE can be a useful measurement in that analysis.

This work is primarily focused on a data center without an economizer in the cooling system. Economizer use and the impact on energy efficiency with regard to data center temperature are discussed later in the results section.

Power consumption in the space

The entire energy chain in the data center needs to be evaluated. Patterson [8] provides a pareto of energy use in the Data center and this can serve as a starting point for the analysis. Each item needs to be evaluated for the impact a different temperature would have upon that component's efficiency and energy consumption.

Basic Setup

For the purposes of the analysis a typical Data Center configuration is assumed. This includes air-cooled CPUs in air cooled servers. The servers are located in stand-alone racks in a data center with a standard hot-aisle / cold-aisle configuration, on a raised floor. A CRAH drives airflow thru a cooling coil in each CRAH and distributes it under the raised floor. Each CRAH has a fan internal to it. A building chilled water system and pump provides cooling fluid to the CRAH. The chilled water is cooled using a standard industrial chiller and evaporative cooling tower.

RESULTS

Server

The primary power users in the server are [8] the CPU, the memory, fans, spinning media, power supply and the planar losses (voltage regulators, printed circuit board, etc...). The most important of these with regard to temperature sensitivity are the CPU and the server internal cooling fans.

The CPU power use is very dependent on temperature and is a critical design feature. This is evidenced by the number of papers discussing logic design and different implementations to limit hot-spots to reduce thermally induced leakage [9], [10], [11]. The temperature has little effect on the dynamic power of the CPU, but primarily drives the amount of leakage. In previous generations of CPU (e.g. 0.35 – 0.18 micron technologies) the leakage was a small fraction of the total but the continued shrinking geometries in the silicon which make the leakage paths shorter have caused the leakage to increase to the point where it can be as high as 50% [12], [13] of the total power. The challenge for CPU designers now is to keep the leakage at or below the 50% level with the ever decreasing geometries.

The leakage is temperature sensitive, the reported ranges vary. Fallah [12] shows a graph which has a consistent 2% leakage increase per °C temperature increase across multiple process generations. Mukherjee [10] reports a 12°C reduction in hotspot temperature gave a 12% reduction in leakage power. The ITRS roadmap [13] suggests that the junction temperatures for high-end server chips should be around 95°C for current and 90°C for future generations. The reduction is based on the need to further control leakage. This lower junction temperature combined with an increase in room temperature makes the thermal challenge even tougher. Fallah [12] shows, for a hypothetical 0.1um technology die, the difference of 10% in total power between running the chip at a junction temperature of 75°C or at 85°C. As technology moves to

45 nm and beyond these temperature dependent leakage rates can be expected to continue. These leakage power deltas must be considered when trying to determine the effect of raising the room ambient temperature.

Server Fans

The second key component in temperature dependent power variation are the server fans. Nearly all modern day servers have variable speed fans driven by platform cooling demand. The fans can operate over a wide range of power draws depending on the thermal load. For example current 1U server platforms fans can draw from 40 watts at peak load to 8 watts at the low end [14]. These values are conservative with anecdotal data indicating some servers may be even double this power use at peak load.

The two components, CPU power and server cooling fans, are also related operationally in the server. The fan speed control (FSC) algorithm in server thermal design is fairly complex and varies by server manufacturer and server type. However there are some baseline assumptions that can be made to assist in the analysis. Generally the server platform measures a number of different temperatures and adjusts the server fans accordingly to ensure all components are at or below their upper temperature limits. It should be noted that the FSC generally will attempt to (within the limit set out above) run the fans at as low a speed as possible to minimize both noise and fan power. Increasing the platform temperatures, including CPU temperature, will generally increase fan speed to increase the required heat removal.

The heat removed is driven by the convective heat transfer in the server. The base equation (considering only the CPU for now) is

$$q = hA_{eff}(T_{heatsink} - T_{air})$$

Where h is the convective heat transfer coefficient and A_{eff} is an area term which also includes the information about the heat sink itself. In this case q must be equal to the heat dissipated by the CPU.

The convective heat transfer coefficient is based upon the Nusselt number and is directly related.

$$Nu = \frac{hL}{k}$$

The Nu is related to velocity in the following form [15]:

$$Nu = a Re^n Pr^m$$

With n ranging from $1/2$ in laminar flow to $4/5$ in turbulent flow. The Reynolds number is linearly related to velocity (V) thru:

$$Re = \frac{\rho V d}{\mu}$$

The convective heat transfer analysis can be complex but in this case we are looking at a fixed geometry in a narrow enough temperature range that we can assume constant properties. Also the forced convection in the server is typically turbulent so we can state that generally

$$h \propto V^{4/5}$$

So to increase the convective heat transfer coefficient the velocity must increase at an even greater rate.

As the temperature in the room increases the available differential temperature decreases.

$$\Delta T = T_{heatsink} - T_{air}$$

$T_{heatsink}$ is fixed by the conduction from the CPU ($T_{junction}$) thru the silicon, the heat spreader (T_{case}), and then the heat sink. With q fixed by the CPU power dissipation the convective heat transfer coefficient must go up by one over the ΔT .

$$h = \frac{q}{\Delta T}$$

The implications of this relationship are significant. As discussed earlier, leakage power is a strong function of temperature. And because of this junction (and case) temperatures are under constant pressure to be set lower to keep the CPU below target design power and leakage. As ΔT gets smaller h must become very large. This can only be done with an even greater increase in velocity (given a fixed server design). And to drive the velocity (flow) higher, the fan speed must increase. The fan laws tell us [16] that this increase is first order, but power consumption actually increases to the 3rd power.

$$\frac{flow_1}{flow_2} = \frac{rpm_1}{rpm_2} = \left(\frac{power_1}{power_2} \right)^3$$

This can represent a significant increase in platform power. Fan power is the largest temperature dependent power use in the platform.

Memory

Memory has become one of the major components of power consumption in the server and the most challenging thermal component in server thermal design. But in this analysis it is not a significant factor.

Memory silicon structure differs from the CPU structure. It is outside the scope of this paper to fully explore those

differences but the key difference is that memory transistors need to switch at a much slower frequency than the CPU allowing the structure to have additional leakage reduction features to be designed in. In addition the sub-threshold leakage (the most temperature dependent and predominant in CPU transistors) is much less a factor in memory than other types of leakage. In general the leakage power will go up but its effect can be considered negligible in comparison to fan and CPU leakage power effects and may be discounted.

Platform components

The power supply unit and the planar losses are only marginally affected by the temperature of the inlet ambient. In the power supply, there are [14] components that increase power usage and some that decrease power usage and as such will generally not significantly affect the total power consumption of the device. Further, the impact is also dependent on % load with lightly loaded devices becoming slightly less efficient (~0.1% /°C) at warmer temperatures and heavily loaded devices becoming slightly more efficient at warmer temperatures (~-0.1% /°C). In either case the value is small compared to CPU leakage and fan power.

Hard drives use a range of power based upon their operational state. When the media is spinning the drive uses more power. When the room is warmer, the drive actually may use less power because of the warming of the grease (lower viscosity) supporting the rotating media. The difference between the states of spinning and idle is on the order of single watts. It is not expected that the variability of the resistance of the grease will vary enough to have an impact in the overall analysis.

CPU/Fan Interaction

Almost all modern servers have variable speed cooling fans internal to the chassis. Fan speed control algorithms are designed to minimize fan power and fan noise by running the fans as slow as possible while keeping all components below their allowed temperature specification [17]. The component temperatures upper limits are all fixed. If the server is exposed to a warmer temperature the driving temperature differential drops, and the fans will run faster. Unfortunately the fan speed control algorithms and thermal capabilities of the server are so varied server to server that we can not draw a simple conclusion about the impact of temperature. It is informative though to consider the limiting cases. The first is the case of a constant fan power over a given temperature range. Assuming in the range of temperatures of interest the fan speed remains constant, then the primary temperature effect in the server will be an increase in CPU leakage. This will increase the heat rejected into the room.

The other limit is one of fixed CPU temperature and variable speed fan control. As the CPU power increases the fan speeds up to maintain the silicon temperature. In this

case the platform power increase is purely fan power. This case assumes that the heat removal is linear with fan speed.

It is unlikely that any platform will show both the full CPU leakage increase and the full fan speed power increase associated with a given temperature rise. The more likely scenario is one close to the second limit with a smaller addition of silicon leakage, primarily due to the fact that the fan speed control will generally have more input parameters than CPU temperature alone and it can be expected that the CPU will be somewhat warmer as the inlet temperature increases. Table 1) displays these two limits and the more likely mixed case.

	CPU Temp (Leakage)	Fan Speed
Limit A	Varies	Fixed
Limit B	Fixed	Varies
Likely	Minor variation	Varies

Table 1) Server power increase cases as a function of room ambient.

Power Delivery Equipment (PDU & UPS)

Power delivery equipment is relatively insensitive to temperature variation within the range discussed. As in the power supply units in the servers, the electrical gear has numerous different components, some which work more efficiently and some that are worse, so the net affect again may be neglected. If the goal is to push into much higher temperatures with equipment that is designed for that purpose the specific efficiencies of the alternate equipment set must be considered.

Transmission

Resistance for copper losses is roughly 1% more for 3°C warmer, but power transmission losses are a low percentage of the total losses and this factor need be considered for only the most precise calculations.

Cooling System

The cooling system is made up of three primary components. The chiller and heat rejection, the chilled water or refrigerant pump and loop, and the air movers and cooling coils in the data center, typically referred to as CRAHs. The cooling system architecture can vary widely and these components may be located closer to or further away from the IT equipment, but the general concept holds for the majority of cooling systems.

Fans/CRAHs

Warmer room temperatures result in warmer air being moved through the room by the fans. This air is easier to move due to the lower density thereby reducing power to move it, but it also has a lower cooling capacity. The Prandtl number is reduced at warmer temperatures, so to

keep things truly equivalent a higher volume of air would need to be transported thru the space to cool at the same capacity. The variations are small (density ~0.3% per °C and Prandtl number even less [15]) and most control schemes do not react to these second order effects so no significant change to the energy use in the fan system is expected.

Chilled water loop

The chilled water system analysis would follow the same logic as the room airflow. The overall impact on room efficiency is small.

Chiller Plant

The primary benefit of the warmer room temperatures to the cooling system is in the chiller plant. In the server the warmer temperature was a challenge to thermal solution because of the reduced differential temperature. In the chiller plant, the efficiency is increased because of an increased differential temperature. The higher the temperature above the outdoor ambient (assumed final heat sink) the more efficient the chiller plant can operate as it is easier to reject the heat with the greater driving temperature difference. The magnitude of the improvement is dependent on the type of chiller, but the Chiller COP can be improved by roughly 1-2.5% by increasing chilled water temperatures 1 deg F. [18]. Raising the room temperature will allow the chilled water temperature to be raised. The total savings will also depend on the percentage of the total power that is represented by the cooling system. Figure 1 shows an average of 35% of the total load is cooling. Only a portion of that is attributable to the chiller. The rest is the room fans (CRAHs) and other cooling system components (e.g. chilled water system)

Alternate Data Center cooling configuration

Air cooled evaporators are common in many mid-sized data centers as the heat rejection component of the cooling system. They are generally less efficient than liquid cooled evaporators and cooling towers, leading to a higher PUE. Because of this larger portion of the energy picture being in the cooling system, these data centers are likely to benefit marginally from a warmer setpoint than a cooling tower based system; however the full analysis of the entire data center infrastructure and IT load is still needed to ascertain the true impact.

Example

Analyzing a conceptual data center can be illustrative in understanding the effect the different parameters have on the efficiency. The values picked are “typical”. A wide range of values could be used as the range of equipment and data center design is quite wide. As an example consider a single rack of forty 1U servers, each with two 100W processors (with leakage at 50W each). These servers each can nominally draw 400 watts. The rack draws 16kW. The data center has a PUE of 2.0 with the cooling portion of the multiplier being roughly 35% and the chiller being roughly 1/2 of the cooling load.

The total data center power will be 32 kW, with 11.2 kW of cooling system power use. The chiller is 5.6 kW of power use. Therefore the overall COP of the chiller system is roughly 4.7 (heat removed/power in).

We will evaluate a change from 20°C (lowest temperature in the ASHRAE recommended range) to 30°C (between the top of the recommended range, 25°C and the maximum allowable 32°C). The 10°C rise will increase the COP of the chiller system by 36%. (2%/deg F) [18]. The chiller system COP would be 6.4. This would improve the PUE to 1.91. There will be an energy savings from the increase in COP.

But there will also be an increase in server power use. We will examine the two limits again. For the case of silicon leakage we will use the value of 2%/°C stated earlier. If the 100W processors draw 100W at 32°C inlet, they will draw less at lower temperatures. The 50W of leakage at 32°C is reduced to 48W at 30°C and 38W at 20°C. With each dual processor server there is a 20W CPU leakage power penalty. We can also assume a voltage regulation efficiency of 85% and a power supply (PSU) efficiency of 80%. Therefore the power increase per server is 29.4 watts. The server rack is now drawing 17.2 kW. With our improved PUE of 1.91 our total power for the rack and overhead is now 32.8 kW.

Now consider the other limit, CPU temperature constant but fan power increasing. If we assume the server used 40 watts at 32°C and 8 watts at 20°C and we apply the fan laws with power changes being related to flow changes to the 3rd power we will use roughly 25.5 watts at 30°C. The fan power will increase 17.5 watts per platform. Considering the PSU at 80% again, the platform will increase 21.9 watts. The rack is now drawing 16.9 kW. The total power consumed is now 32.2 kW.

	T _{amb}	CPU Leak	Fan Pwr	PUE	Rack Pwr
Baseline	20°C	low	slow	2.0	32 kW
Limit A	30°C	varies	fixed	1.91	32.8
Limit B	30°C	fixed	varies	1.91	32.2

Table 2) Room power increases at warmer temperatures

In both of the limiting cases (a. constant fan power with variable leakage, b. constant CPU leakage & temperature with variable fan power) the total power consumed went up but the goal was to improve chiller efficiency and reduce total power. Even with a 36% improvement in Chiller system COP the additional power consumed in the servers at the warmer temperatures outweighed the cooling system improvement.

We can now illustrate the impact of an airside economizer. We will assume the chiller is shut off making the COP of the chiller system infinite. The PUE of this data center is

now 1.65 (2.0 – 0.35). Considering our more likely case above of variable fan speed the total power consumed drops from 32kW to 27.9kW.

	T _{amb}	CPU Leak	Fan Pwr	PUE	Rack Pwr
Baseline	20°C	low	slow	2.0	32 kW
Economizer	30°C	fixed	varies	1.65	27.9

Table 3) Power savings with economizer cooling system

The preceding analysis used assumed numbers but when compared to Figure 1 a PUE of 2.0 and a cooling load factor of 35% fit the benchmarking data very nicely. For the increased chiller efficiency to outweigh the added power added by the servers being in a warmer environment would require a PUE much greater than 2.0. In those cases there are likely other opportunities in the Data Center to reduce energy consumption with lower risk than increasing the ambient set point.

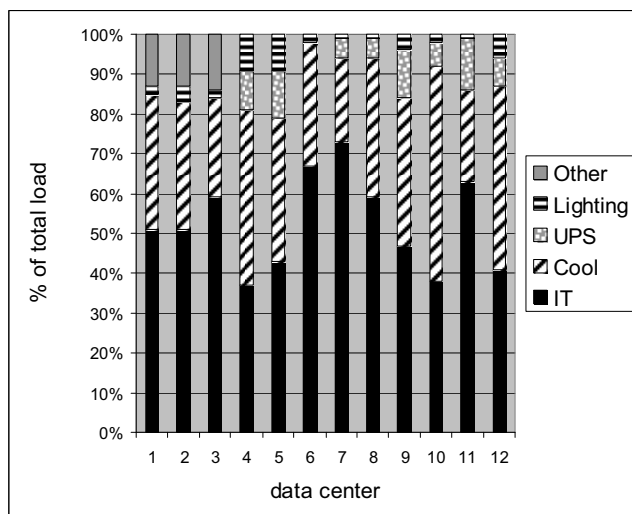


Figure 1 Data Center Benchmarking (LBNL, [3]), reformatted

The analysis could easily be repeated for different temperature increases, of better PSU efficiencies, or lower leakage parts (e.g. older CPUs leaked less) but the end result will be similar, there are no significant savings to the non-economizer based data center by merely increasing ambient temperature. Only with an economizer are the energy savings realized.

Other impacts

Beyond the implications of the mixed results for energy savings, there are other issues related to increasing ambient temperatures. The first is reliability. The main physical models for reliability generally have temperature as a parameter. There is no single relationship that correctly correlates temperature and reliability as each component in the server is affected differently. However a number of

specific reliability issues are noted. An 8°C temperature increase is reported to reduce battery life by 50 percent. Similarly a 10°C rise in a capacitor component temperature will reduce its life by ½. There can be reliability impacts and these need to be weighed.

Room impacts

There other impacts to the room and these need to be considered as well. First are the ergonomic issues of the warmer room. While the cold aisles will generally be a suitable temperature, a well designed data center with hot aisle segregation may result in hot aisle temperatures, that when combined with warmer room ambient temperatures are warmer than ideal as a working environment and could require a detailed ergonomics and safety review. The other potential impact is acoustics. As with fan power, the increased fan speed will generate even greater noise. Power went up roughly to the 3rd power, noise generally increases to the 5th power. The higher acoustic levels in the warmer room may require hearing protection and an increased acoustics management plan.

Airflow balance

Server fan increases may put the airflow in the room out of balance. Initially we assumed the CRAH had enough airflow to provide the appropriate volume of cool air to each server inlet. As the room temperature is increased the server fans are likely to speed up. If the CRAH can no longer supply the volume of ambient air needed and the server and rack begins to recirculate hot aisle air into the server inlet the airflow balance in the room can go non-linear quickly with the undesired result of server throttling being the end state.

RECOMMENDATIONS

Assuming the room is currently operating with-in the ASHRAE recommended range (20-25°C), pursuit of a warmer ambient strategy should be explored cautiously. It would be a reasonable first step to modulate the room set point some small but measurable amount (e.g. 2°C) and measure total power consumption, including all IT equipment as well as utility infrastructure. In addition, any other operational changes should be noted, including supply air temperatures to all the servers. The set point temperature variations should be both higher and lower (e.g. if the room is typically at 23°C, run it at 21°C and 25°C). This will yield a difference twice as large in terms of measuring power savings for minimal risk to stable operations. If no savings are noted it is unlikely more will be found at warmer room temperatures. It should also be noted that the measured differences will likely not be linear as the room temperature is increased due to the power law growth rate of server fan power. Each new room setpoint should be approached gradually, again with measurements of all key systems and components to understand the overall impact to efficiency.

In the case of room temperatures below the ASHRAE recommended range (less than 20°C) raising the room

temperature should be pursued more aggressively as there is no benefit to running the data center at that cool a temperature, and it is likely using more energy than is needed. Additionally, the increases in fan power or CPU leakage in increasing room temperatures that are below the ASHRAE recommended values will be minor. But even in this case, the precautions laid out above should be considered in increasing the room ambient.

If the room is cooled with a cooling system that includes an economizer the analysis shifts significantly. Raising the temperature to the upper limit of the ASHRAE recommended range or even a couple degrees beyond can allow the use of an economizer. With room setpoints at the lower end of the recommended range or even below, the hours per year of potential economizer use is often not enough to have a positive ROI and often they do not get installed. If the data center were designed at 25°C or 27°C the number of locations and that would have a positive ROI for an economizer installation greatly increases based on the number of hours the system could be run with the chiller off. This is the true advantage of raising room temperatures; increasing economizer use and getting rid of the power demands of the chiller for more of the year.

Unfortunately obtaining this positive ROI is much simpler in the case of new construction than a retrofit. The complexity of adding the economizer to an existing facility will preclude it from installation in any significant numbers of existing data centers.

CONCLUSIONS

From the simplest thermodynamic evaluation the thermal solution in the data center must move heat from the ultimate source, (in the simplest analysis the CPU at T_{junction}), all the way to the ultimate heat sink the outdoor ambient environment (at T_{outdoors}). Changing an intermediate temperature in that path (T_{air}) does not change the endpoints and may or may not improve the efficiency. What has been demonstrated is that as the data center temperature is raised the overall amount of heat that has to be rejected to the final heat sink actually increases. The efficiency of a major link in the thermal management system (specifically the chiller) can be improved, but the overall energy consumed can actually go up.

On the other hand, changing the thermal management system to frequently replace the chiller with an economizer which can run for a significant portion of the year due to a warmer room setpoint is the most promising method of reducing energy costs associated with data center cooling.

References

[1] ASHRAE, *Thermal Guidelines for Data Processing Environments*, 1st Edition, ASHRAE, Atlanta, 2004
 [2] The Green Grid, "The Green Grid Power Efficiency metrics; PUE and DCiE", www.thegreengrid.org, 2007.

[3] Lawrence Berkeley National Labs, Benchmarking: Data Centers, Dec 2007, <http://hightech.lbl.gov/benchmarking-dc-charts.html>

[4] California Energy Commission http://www.consumerenergycenter.org/tips/business_summary.html, Feb 2007

[5] <http://www.processor.com>, "Going Green In The Data Center, Practical Steps For Your SME To Become More Environmentally Friendly" September 28, 2007 • Vol.29 Issue 39 Page(s)

[6] http://www.epa.gov/cleanrgy/pdf/sector_meeting27_Jun07/4bi_officebuilding.pdf

[7] Patterson, M.K., Costello, D., Grimm P, Loeffler, M. (2007) "Data Center TCO; A Comparison of High-density and Low-density Spaces," THERMES 2007, Santa Fe, NM

[8] Patterson, M.K., Pratt, A., Kumar, P., "From UPS to Silicon, an End-to-End Evaluation of Data Center Efficiency, Proceedings of the EPA Event: Enterprise Servers and Data Centers: Opportunities for Energy Savings." (February 2006)

[9] Ni, Min and Memik SO, "Thermal Induced Leakage Power Optimization by Redundant Resource Allocation", ICCAD 2006, Nov 5-9, 2006, pp297-302.

[10] Mukherjee, R, et.al. "Peak Temperature Control and Leakage Reduction During Binding in High Level Synthesis", ISLPED '05, Aug 8-10, 2005

[11] Agarwal, A., et. al. "Leakage power analysis and reduction: models, estimation, and tools", IEE Proceedings, 2005

[12] Fallah, F. and Pedram, M., "Standby and Active Leakage Current Control and Minimization in CMOS VLSI Circuits", IEICE Transactions on Electronics, 2005, Vol E88-C, Num 4, pp.509-519

[13] International Technology Roadmap for Semiconductors, 2006 Update http://www.itrs.net/Links/2006Update/FinalToPost/11_AP2_006UPDATE.pdf

[14] B. Griffith, Intel Corporation, personal communication, December, 2007

[15] Incropera, F. P. and Dewitt, D. P., *Fundamentals of Heat and Mass Transfer*, 5th Edition, John Wiley and Sons, 2002

[16] Buffalo Forge Company, *Fan Engineering*, Edited by Robert Jorgensen, Eight Edition, Buffalo, 1983

[17] Patterson, M. K., Steinbrecher, R. and Montgomery, S. "Data Centers: Comparing Data Center & Computer Thermal Design", ASHRAE Journal, Vol 47, No 4., April 2005

[18] Dubin, Fred S. and Long C., *Energy Conservation Standards for Building Design Construction and Operation*, 1st Edition, McGraw Hill, 1978