An Approach to Calculating and Defining Overall Data Center Energy Efficiency including Compute, Network, Storage and Facilities

Ed Ansett (Singapore) & James Dow (New York) - i_3 Solutions Group

Abstract

The definition of data center efficiency in terms of the facility elements is well established. The universally adopted metric, Power Usage Effectiveness (PUE) [1] has several categories.

There are two important issues associated with PUE.

1. PUE is not intended to take into account the energy efficiency of the IT equipment within the data center.

2. There are emerging trends toward locating some power and cooling functions that are normally part of the facility within the IT equipment. This obscures the demarcation between the facility and IT and can skew PUE results.

In this paper a suggested framework for defining and calculating IT energy efficiency in terms of compute, network and storage in conjunction with PUE is discussed. The combination of these metrics results in a new metric for the overall data center efficiency.

The principle that underlies the approach taken is predicated on knowledge of the energy efficiency of a particular device (server, network node or storage) for the operating conditions that the particular device will experience. In this context, energy efficiency is considered up to the point, but not including, the point at which data state change occurs.

It is acknowledged that currently the gathering of data required to achieve this goal is not straightforward particularly for network and storage components. However most of the data already exists for compute in the form of OEM calculators.

1. Introduction

A holistic understanding of data center efficiency should take account of:

1. Compute environment workload
2. Data flows in networks
3. Data density and I/O rate in storage
4. Data center facility loads

To be consistent with the approach taken with the PUE definition, determination of data center energy efficiency should not take into account system performance issues. Consider for example the PUE of a data center such as the Uptime Institute Tier 3 type and that of a Tier 1 type data center [6]. Inevitably the higher availability system yields a less energy efficient PUE all other things being equal due to the additional components and their associated fixed losses.

There has been considerable debate regarding what IT output variables should be included in a definition of IT energy performance. For example in terms of compute; should this be the number of instructions, transactions, clock cycles per unit of power? Analogous issues exist for networks and storage. For network should it be bit rate, packet rate, per unit of power? For storage: stored bytes, I/O rate per unit power etc. Earlier work [10] provides insight into how overall data center energy performance could be addressed.

In this alternative approach of overall IT energy efficiency, consideration of IT energy loss is limited to any point at which power is transmitted or converted prior to delivery to a consuming IT component that performs data state change. This includes losses due to internal switched mode power supplies, internal cooling fans and DC-DC converter losses off the backplane. Losses associated with the internal IT components in performing state change, e.g. the efficiency with which power is converted to processor cycles within the processor, are out of scope. The overall IT efficiency definition proposed in this paper considers the energy efficiency of the components shown in Figure 1.

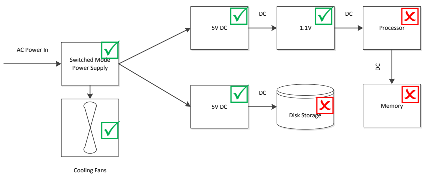

Figure 1

The result of this approach is that performance variables are no longer relevant unless they have an impact on the energy efficiency of the data center. Therefore data center performance in terms of desired outputs other than energy efficiency is beyond the scope of this paper. The overall data center efficiency definition proposed in this paper considers the energy efficiency of the facility and IT systems shown in Figure 2.

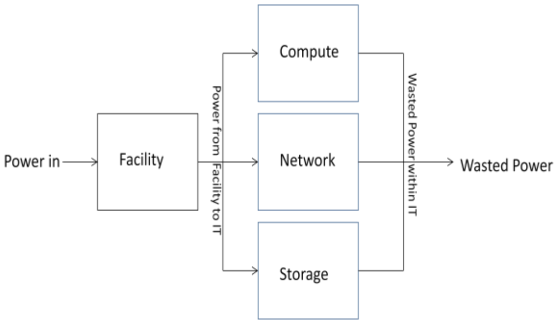

Figure 2

This is conceptually the product of facility energy efficiency and the sum of individual energy efficiencies of compute, network and storage each weighted according to its relative energy consumption to overall IT energy consumption (1)

$$\eta_{DC} = \frac{(k_c\eta_c + k_n\eta_n + k_s\eta_s)}{PUE} \qquad (1)$$

Where:

η_{dc} is the total data center energy efficiency

PUE = Power Usage Effectiveness

k_c = relative power of compute to the overall IT power

k_n = relative power of network to the overall IT power

k_s = relative power of storage to the overall IT power

η_c = energy efficiency of compute devices

η_n = energy efficiency of the network

η_s = energy efficiency of storage devices

2. Facility and IT Delineation for PUE Calculation

PUE becomes problematic when power or cooling functions that are usually part of the facility infrastructure are located within the IT infrastructure.

For example in an enclosed hot or cold aisle containment system replacing the computer room air handling units (CRAHs) with overhead cooling coils and utilizing the IT equipment internal fans to provide all the energy for air movement within the aisles and through the equipment racks; thus energy that used to be part of the facility side of the PUE calculation is now part of the IT load in the PUE calculation.

A similar issue exists with the replacement of the conventional UPS within the facility by a battery and charger integral to the server.

It is evident from these and other examples that the delineation between where the facility ends and the IT infrastructure begins has become increasingly unclear. Figure 3 refers.

Figure 3

The implementation of virtualized server environments can cause a worsening of PUE [2] even though the data center energy overall is reduced i.e. an improvement in overall data center efficiency can lead to the degradation of PUE when virtualization is applied and the facility base load remains constant. To maintain PUE the facility infrastructure capacity must be reduced otherwise virtualization can result in facility power and cooling over-provisioning.

3. Compute Useful Work

Recently a definition of Useful Work in the context of the compute environment defined compute efficiency in terms of the amount of primary services provided by the server [3]. Primary services are considered useful work and exclude overheads i.e. secondary and tertiary services such as anti-virus, drivers, and software firewalls etc.

The metric definition for server compute efficiency (ScE) is percentage given by (2):

$$ScE = \frac{\sum_{i=1}^{n} p_i}{n} \times 100 \qquad (2)$$

Where n is the total number of samples and p_i is the number of samples providing primary services. Applying this metric to the whole data center comprising m servers gives the data center compute efficiency metric (DCcE) given by (3):

$$DCcE = \frac{\sum_{j=1}^{m} ScE_j}{m} \times 100 \qquad (3)$$

Whilst this is a useful measure of server and data center useful work in terms of compute output it does not describe the energy efficiency of the server or the data center.

4. Compute Energy Efficiency

Compute energy efficiency has increased substantially in recent years primarily due to virtualization of servers and the introduction of much improved switched mode power supply efficiencies. Other significant improvements in compute energy efficiency include the adoption of server energy management tools, power capping and varying CPU clock speed.

In the early 2000's typical server power supply efficiency at part load (say 50%) was typically 60% and exhibited poor correlation between CPU activity and energy efficiency. In the last few years contemporary high efficiency servers, network switches and routers, and storage devices have internal switched mode power supplies that operate at their highest energy efficiency at around half electrical load. Figure 4 refers.

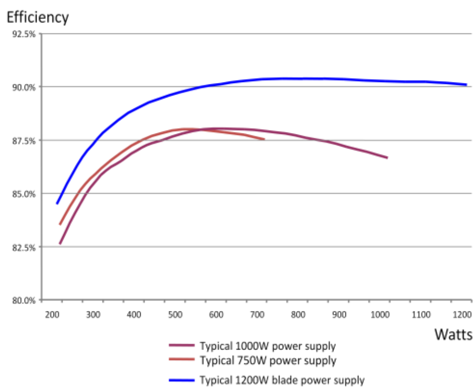

Figure 4

Power supplies are designed this way by hardware manufacturers because this is the point where the equipment is expected to operate most of the time. Yet despite this for servers operating in normal energy mode the quiescent power at 0% CPU utilization is excess of 50% of the server full load power. Figures 5 and 6 refer.

This factor is a challenge for manufacturers because it represents a substantial unproductive energy overhead in server environments. One response to this is to utilize sleep-states e.g. Intel C and DC sleep-states; however there is a trade-off in terms time required to wake from the sleep states.

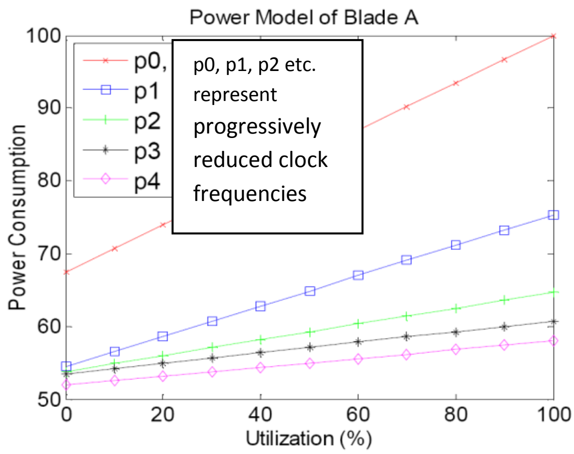

Figure5

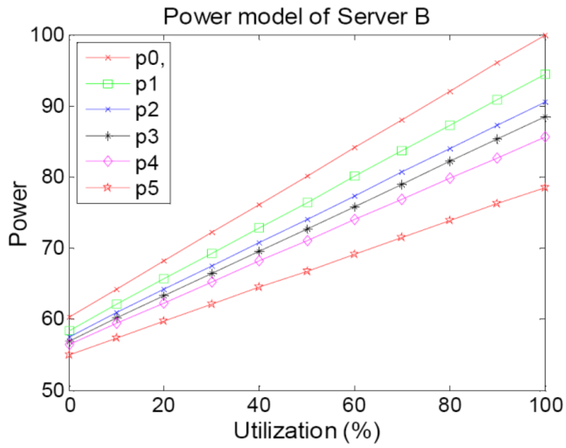

Figure 6

In a virtualized server environment the objective is to make better utilization of compute resources and reduce energy consumption by combining multiple virtual machines onto a lesser number of physical servers. Server virtualization has provided a significant increase in overall data center energy efficiency [2]. It is reasonable therefore that any new measure for server energy efficiency should includes the impact of virtualization.

The typical relationship between CPU activity and power consumed for mid-range blade and conventional flat chassis servers is illustrated in figures 5 and 6.

Several points can be noted.

1. At zero utilization the servers require in excess of 50% of their full load power

2. The relationship between energy consumption and CPU utilization is (usually close to) a linear straight line

3. Reducing clock speed to effect energy savings is more effective as CPU utilization approaches 100%

Whilst the relationship between server power consumption and CPU utilization generally follows a linear straight line, the relationship between power consumption and energy efficiency is non-linear as illustrated in Figure 4.

For enterprise class servers both CPU (and GPU) utilization has become less dominant from an energy standpoint primarily due to multi-core processors. The additional processing capacity has brought both memory and I/O capacity into play as factors driving energy consumption and efficiency in conjunction with CPU / GPU utilization.

Whilst it may be possible to find a linear equation that is a best fit to the various energy efficiency curves generally a more accurate approach is to use a computer simulation using static arrays to convey the server energy efficiency at discrete intervals and interpolate the energy efficiency [9].

To determine server energy efficiency consider a static array that contains the server energy efficiency data η at operating points: k CPU utilization, p memory configuration and r I/O configuration.

Then energy efficiency can be interpolated from this data. And the average server energy efficiency (η_{server}) is given by (4):

$$\eta_{server} = \frac{\sum_{m=1}^{M} \eta_{(k,p,r)}}{M} \quad (4)$$

Where:

$\eta_{(k,p,r)}$ is the server energy efficiency at CPU (GPU) utilization rate k, memory configuration p and I/O configuration r.

m is the sample number and $(1 \leq m \leq M)$

M is the total number of samples

Then for N servers in the data center the average energy efficiency of compute is given by (5):

$$\eta_c = \frac{\sum_{n=1}^{N} \eta_{server}}{N} \quad (5)$$

Generally for servers the required data already exists in the form of energy calculators provided by hardware manufacturers that enable the end user to know the relative power consumption from 0-100% utilization for most server configurations. Examples include: IBM Systems Energy Estimator, HP Power Advisor, Cisco Power Calculator, Oracle Calculator, Dell ESSA and Fujitsu System Architect.

5. Network Energy Efficiency

Work to improve the energy efficiency of networks has also made significant progress in recent years.

Network equipment tends to use high efficiency switched mode supplies as illustrated by the Cisco Catalyst 6500 in Figure 7.

Other approaches to reduce power consumed by the network using techniques such as Energy Aware Routing in the Cognitive Packet Network [7] and Reduced Network Energy Consumption via Sleeping and Rate-Adaptation [8].

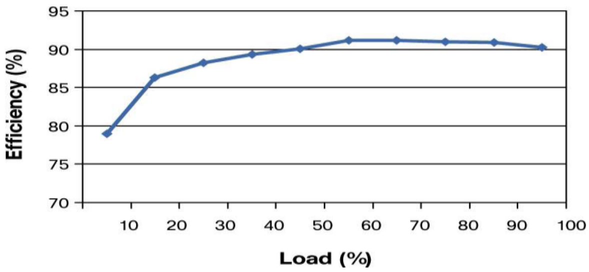

Figure 7

Network equipment does not always exhibit an obvious correlation between network activity and electrical load for the following reasons.

1. Routers and switches may have multiple processors to manage traffic, not just a central processor. There may be multiple line cards, a management processor, and traffic management components.

2. If the equipment is providing Power over Ethernet (POE) there are power losses in the Ethernet cabling serving downstream devices and the downstream devices themselves.

3. Energy efficiency initiatives that switch power distribution across the network

switch stack, such as Cisco Stackpower [4], allows switch devices to share power and improve reliability with a common distributed power system.

4. Each type of node inter-connect methodology, i.e., fiber versus copper, 1Gb/s versus 10Gb/s, has a unique coefficient of efficiency.

Network reliability and Quality of Service (QoS) are outside the scope of this paper [5]

Recent work [9] to profile to power consumption of networks in terms of network node workload using a Power Consumption Profiling System under various operating scenarios in a Multiprotocol Label Switching (MPLS) network suggests that power consumption in a network node is driven by packet processing rate rather than bit rate. This is because CPU activity is driven by processing packet headers, rather than processing the body of data within the packet.

Irrespective of the type and complexity of the network, if the network node power efficiencies are known at discrete intervals for each packet size and packet rate then interpolation can be used to determine node power efficiency for any operating condition.

Consider a static array which stores the network node energy efficiency η at operating points *k* packet rate, *p* packet size and *c* interface type.

Then energy efficiency can be interpolated from this data. And the average network node energy efficiency (η_{node}) is given by (6).

$$\eta_{\,node} = \frac{\sum_{m=1}^{M} \eta_{\,(k,p,c)}}{M} \qquad (6)$$

Where:

$\eta_{(k,p,c)}$ is node energy efficiency at packet rate *k* and packet size *p* with interface type *c*.

m is the sample number (1 ≤ m ≤ M)

M is the number of samples

Then for N nodes in the average data center the network energy efficiency (η_n) is given by (7)

$$\eta_{\,n} = \frac{\sum_{n=1}^{N} \eta_{\,node}}{N} \qquad (7)$$

6. Storage Energy Efficiency

Energy consumption in storage systems is a function of the capacity of the storage device, number of disk drives and I/O rate, as these are the components that drive the variability of consumption.

Numerous measures have been introduced by storage manufacturers to reduce energy consumption that frequently improves resource utilization; e.g. storage virtualization, auto-tiering, lower power disks, and thin provisioning. It is significant to note that these measures primarily address the efficiency of the IT data state change, and are thus beyond the scope of this analysis.

Other measures introduced to drive efficiency, e.g., high efficiency switched mode power supplies, high efficiency disk actuators, etc., do have impact on the overall energy efficiency prior to data state change.

The approach to determine storage system energy efficiency follows is as follows.

Consider a static array that contains the storage device energy efficiency η at operating points *k* storage capacity and *p* I/O rates.

Then given the various operating conditions of the storage device the average energy efficiency is given by (8).

$$\eta_{\,array} = \frac{\sum_{m=1}^{M} \eta_{\,(k,p)}}{M} \qquad (8)$$

Where:

$\eta_{(k,p)}$ is array energy efficiency at storage capacity *k* and I/O rate *p*.

m is the sample number

M is the number of samples (1 ≤ m ≤ M)

Then for N arrays in the data center the average storage energy efficiency (η_S) is given by (9)

$$\eta_{\,s} = \frac{\sum_{n=1}^{N} \eta_{\,array}}{N} \qquad (9)$$

Overall Data Center Energy Efficiency Equation

With the energy efficiency determined for compute (η_c), network (η_n) and storage (η_s) the overall energy efficiency of the IT infrastructure (η_{IT}) can be determined based upon their respective power allocation. Consider the power and cooling load factor allocation across the entire data center as follows: k_c to compute, k_n to network and k_s to storage such that $k_c + k_n + k_s = 1$ then:

$$\eta_{IT} = k_c\eta_c + k_n\eta_n + k_s\eta_s \qquad (10)$$

And the overall energy efficiency of the data center (η_{DC}) is given by (11):

$$\eta_{DC} = \frac{\eta_{IT}}{PUE} \qquad (11)$$

Conclusion

In this paper we have addressed the issue of defining a single metric for data center energy efficiency including compute, network, storage and facility elements. In the process of doing this we have introduced definitions for compute, network and storage energy efficiency.

The definitions reflect the energy efficiency of the various systems under consideration in the context of their workload considering only the electrical power chain and internal cooling fans, not the efficiency of the data state change. This can be reflected as an overall average static number for a given period of time or can be calculated in real time.

The new metric also obviates the issue of demarcation between facilities and IT by taking account of the energy losses that moved from facility to IT.

For compute the variable that drives energy efficiency is CPU utilization, memory and I/O configuration. For network it is packet size, packet rate and interface type. For storage it is storage capacity and I/O rate.

In order for the proposed metric to be used accurately in practical situations IT hardware manufacturers are requested to provide additional energy efficiency calculators for their devices in a format that is consistent across all vendors.

In a server context the data required for static array population is currently available; albeit not in a consistent format.

Currently there are not enough energy efficiency data sets available from storage and network vendors.

It may be necessary to add the thermal operating condition to one or more of the data sets. This and indeed all energy efficiency drivers are matters for the manufacturers to confirm and if required include in the data sets.

7. Next Steps

Further development to practically implement this approach to calculate the overall energy efficiency of the data center (η_{DC}) will require effort on the part of management software companies to extract the operating condition for each IT device and calculate the efficiencies by interpolating data sets containing operating condition and energy efficiency for compute (η_c), network (η_n) and storage (η_s).

In addition the hardware manufacturers will need to make the data sets available in a consistent format for the various operating conditions and the associated energy efficiency at each discrete level.

It is likely that the variables that drive energy efficiency may change in future. A recent example of this type of change is the hitherto dominance of CPU utilization in enterprise servers. This has diminished with the advent of multi-core processing. This in turn resulted in increased significance of memory and I/O with regard to server energy efficiency.

The approach outlined in this paper will need to adapt over time to changes in technology energy drivers. However the general methodology of using interpolation of efficiency data sets based on discrete operating conditions (provided by the IT hardware OEM community), and data extraction and analysis (provided by the management software community) will provide data center owners and operators real time knowledge of data center energy efficiency for the foreseeable future.

Acknowledgments

Ken Baker, Hewlett Packard

Robert Tozer & Sophia Flucker, Operational Intelligence

Ricardo Lent, Imperial College London

Peter Gross, Bloom Energy

References

1. Recommendations for Measuring and Reporting Overall Data Center Efficiency, 2011. 7x24 Exchange, ASHRAE, The Green Grid, Silicon Valley Leadership Group, Save Energy Now, Energy Star, USGBC, The Uptime Institute

2. Virtualization and Cloud Computing: Optimised Power, Cooling and Management...Schneider Electric, Suzanne Niles and Patrick Donovan

3. Data Center Compute Efficiency Metric: The Green Grid. Azevedo, Hawkins et al. 2010

4. Cisco StackPower: Efficient Use of Power. Cisco Systems Inc. 2012

5. Understanding Network Failures In Data Centers. Gill, Jain, Naggapan. August 2011

6. Tier Classifications Define Site Infrastructure Performance. Turner, Seader, Renaud, Brill. 2008

7. Energy Aware Routing in the Cognitive Packet Network. Gelenbe and Mahmoodi. 2011

8. Reduced Network Energy Consumption via Sleeping and Rate-Adaptation. Nedevschi et al.

9. Simulating the Power Consumption of Computer Networks, Lent, 2010 IEEE International Workshop on CAMAD

10. Japan’s Approach to reducing Greenhouse Gas Emissions from Data Centers, Shino, 2010 Nomura Research Institute

Example Calculation of η_{DC}

Consider a 1 MW data center with a known PUE of 1.35 (i.e. the facility is 74.07% efficient in terms of the former DCiE metric). This means that 740.07kW is the average annual IT load.

Assume that the IT load comprises 55% servers, 15% network and 30% storage. Therefore from Eqn (10):

k_c – compute power and cooling constant = 0.55

k_n – network power and cooling constant = 0.15

k_s – storage power and cooling constant = 0.30

Given the following average energy efficiencies provided by management software comprises

η_c – compute = 0.85

η_n – network = 0.65

η_s – storage = 0.75

Then the overall IT energy efficiency (η_{IT}) is:

$$\eta_{IT} = k_c\eta_c + k_n\eta_n + k_s\eta_s \qquad (10)$$

In this case:

$\eta_{IT} = (0.55 \times 0.85) + (0.15 \times 0.65) + (0.4 \times 0.75)$

$\eta_{IT} = 0.865$ (i.e. 86.5%)

And the overall data center efficiency (η_{DC}) is:

$$\eta_{DC} = \frac{\eta_{IT}}{PUE} = \frac{0.865}{1.35} = 0.64 \qquad \text{i.e. } 64\%$$

Therefore in this example facility energy efficiency is 74% (PUE 1.35) and the overall data center efficiency (η_{DC}) including both PUE and η_{IT} is 64%.