THE GREEN GRID DATA CENTER COMPUTE EFFICIENCY METRIC: DCcE

EDITOR(S):
Mark Blackburn, 1E

CONTRIBUTOR(S):
Dan Azevedo, Symantec
Andy Hawkins, 1E
Zeydy Ortiz, IBM
Roger Tipley, HP
Dr. Sven Van Den Berghe, Fujitsu

 Rev 2010-0

Executive Summary

The Green Grid (TGG) is a global consortium of companies, government agencies, and educational institutions dedicated to advancing energy efficiency in data centers and business computing ecosystems. The Green Grid is proposing the use of a new metric—data center compute efficiency (DCcE)—and its underlying sub-metric, server compute efficiency (ScE).

These metrics will enable data center operators to determine the efficiency of their compute resources, which allows them to identify areas of inefficiency. Using DCcE and ScE, data center operators can continually refine and increase the efficiency of their compute resources in the same way that they use power usage effectiveness (PUE) to improve data center infrastructure.

Reducing energy demand by increasing compute efficiency has a larger impact on overall data center energy use due to the consequent reduction in the power and cooling infrastructure load. The Green Grid's recent study on unused servers[1] revealed a significant percentage of unused compute resource within most data centers. These new compute efficiency metrics will make it easier for data center operators to discover unused servers (both physical and virtual) and then decommission or redeploy them.

DCcE is not a productivity metric; it does not show how much work is being done by the data center. However, it can be used in conjunction with some of the proposed data center productivity (DCP) proxies[2] to determine the proportion of measured work that is providing the primary services of the data center, compared to the amount of energy consumed.

This white paper outlines how to measure DCcE using information that is easily extracted from server operating systems—without needing to re-instrument any applications.

Table of Contents

I. Introduction

Any of the metrics proposed for data center productivity require an understanding of how much useful work is being performed. Yet there has been no easy way to determine because of the difficulty in defining "useful."

When looked at from the perspective of individual microprocessor instructions, every instruction processed obviously has some functionality and could be considered useful. This is how most computer systems treat work—as a percentage of time doing anything other than idling—and they usually refer to it as "utilization."

The success of server virtualization initiatives in improving efficiency has mainly been measured by the increase in overall processor utilization, since a server is using power more efficiently when more of the power is being used to compute rather than just idle. However, CPU utilization only shows how *much* work is being done; there is no determination as to whether the work being done is *useful*.

This paper aims to deliver a mechanism for determining the proportion of work that provides primary services, server compute efficiency (ScE), which can then be aggregated to determine the compute efficiency of the entire data center, data center compute efficiency (DCcE).

ScE (and by extension DCcE) is a time-based metric; it measures the proportion of time the server spent providing primary services and is expressed as a percentage value.

II. What is Meant by Primary Services?

A server (whether physical or virtual) is usually commissioned to provide one or more specific services—to "serve," hence the name. For the purposes of this white paper, these services are referred to as the primary services for that server. Secondary, tertiary, and similar services may be optional or required in order to provide a primary service, but if the primary service did not exist, there would be no reason for the secondary and tertiary services to continue running.

As a generic example, the primary service of an e-mail server is to provide e-mail. This same server may also provide monitoring services, backup services, antivirus services, etc., but those are secondary, tertiary, and similar types of services. If the e-mail server stops being accessed for e-mail, the monitoring, backup, and antivirus services may no longer be necessary, but the server may still continue to provide them. So from a CPU-utilization standpoint, the unused server may appear to be busy, but that may only be secondary or tertiary processing. Hence, CPU utilization is not a precise enough measure.

It therefore logically follows that if the server is doing anything other than the work for which it was commissioned that work is not a primary service.

The processes associated with specific applications on individual servers are normally well known. Plus, server operating systems have a mechanism to provide utilization statistics that are broken down by process. Therefore, it is possible to work out how much of an individual server's utilization is dedicated to the specific application providing the primary service and, from this, to ascertain the proportion of primary services that the server is supporting.

There are several problems with this method:

- Not all servers have well-defined processes performing primary services (e.g., terminal servers do not have well-defined processes since users can be running a myriad of end-user processes).
- In large data centers, tracking and recording which processes on which servers were providing primary services would require a great deal of administrative overhead.
- The method does not take into account the proportion of underlying operating system activity that contributes to the primary services.

III. Primary Services by Subtraction

In any well-run data center, the processes that are providing secondary, tertiary, and other services usually form a relatively small set of well-known maintenance and monitoring tasks, including:

- Anti-virus
- Backup
- Disk defragmentation
- Drivers
- Indexing
- Monitoring
- Network management
- Patch management
- Screensavers
- Setup routines
- Storage management
- System processes
- Software firewalls
- Systems management
- Time synchronization
- Utilities
- Virtualization management

Some of these processes may actually be primary services on some servers, such as those whose purpose is to provide backup, systems management, and so on. In these cases, the server components of the service are usually different from the client processes that provide secondary services on other servers, and therefore the server processes will register as primary services anyway. For example, the processes running on a backup server are different from the backup client agent process that runs on the client. The backup client agent would be defined as a secondary or tertiary service, while the backup server process would appear as a primary service.

The set of secondary/tertiary services in use across the data center is relatively small compared with the set of primary services, and those secondary/tertiary services are likely to be common across servers with different application types and even different operating system versions. Thus, an easier mechanism to ascertain the proportion of utilization attributable to primary services is simply to take all utilization (total CPU utilization) and then subtract from it the utilization that is known to be from secondary, tertiary, and other services—leaving only primary services utilization.

Since there is a certain amount of "autonomic" activity performed by the server operating system itself, it also is necessary to screen out this "noise" by imposing a minimum threshold below which primary services are not considered to be performed.

It should be noted that primary services are not constrained entirely to processor activity. A server that is providing large amounts of input/output (I/O), such as a file server, could be supplying primary services without utilizing the processor much at all. Similarly, a terminal server being used for normal productivity-type applications may use very few processor resources but still provide primary services to its end users. Therefore, it is necessary to use a wider set of methods to determine when primary services are being provided.

IV. Primary Services I/O

I/O can be used the same way as CPU to determine primary services activity—take the total I/O and subtract the I/O caused by secondary, tertiary, and other services, which leaves only the primary services I/O. Also, as with CPU, a threshold can be used to screen out the operating system's own autonomic I/O noise.

For the purposes of ascertaining primary services activity, any and all types of I/O—including network, disk, and device I/O—should be used.

V. Incoming Requests

Since a server has to serve, any client requesting a service must communicate with the server across the network, and it must do so using a session-based protocol (TCP is the most common at present). Incoming communications via connectionless or non-session-based methods must be ignored since those communications can only be broadcast, multicast, or fire-and-forget types of traffic that do not require a dedicated response from a specific server.

Incoming requests for specific services (for example a web page, an SQL query, a DNS lookup, etc.) use TCP sessions to ensure that the request reaches the appropriate server and that the response goes to the correct client and can be understood by that client. Therefore, if any of the processes of a primary service receive an incoming session-based connection request, it can be assumed that primary service activities are being performed even if they result in relatively little I/O or CPU utilization.

VI. Interactive Logons

Terminal server–type applications may not necessarily register when using the methods above to determine primary services activity because they may not create very much CPU activity or I/O and because the incoming remote-access sessions can be long-lived. One can, however, ascertain if a server is providing this type of primary service by examining when interactive logons occur.

VII. Deduction of Primary Services Activity

For the purposes of the ScE metric, all four of the above methods should be used to determine if a server has active primary services.

For each method, measurements should be taken at a regular interval (e.g., once per minute). The server being measured can be deemed as having provided active primary services during that period of time if any of the following criteria are met:

- The average amount of CPU utilization attributable to primary services (total average CPU utilization minus average CPU utilization from secondary and tertiary services) is above a designated threshold, such as 10%, which has proven to be an effective choice.
- The amount of I/O attributable to primary services (total I/O minus I/O from secondary and tertiary services) is above a particular threshold; experience has shown 500Kb/sec to be a good threshold.
- A primary services process (not a secondary or tertiary service) has received an incoming session-based connection request.
- There has been an interactive logon to the server.

If none of the above criteria are met, then the server can be considered as not having provided primary services during that time sample.

The ScE percentage over any time period is therefore calculated by summing the number of samples where the server is found to be providing primary services (p) and dividing this by the total number of samples (n) taken over that time period and multiplying by 100.

$$ScE = \frac{\sum_{i=1}^{n} p_i}{n} \times 100$$

VIII. Sampling Frequency—Accuracy versus Performance Impact

Since ScE is based upon a set of criteria either being met or not, the accuracy of this metric is dependent upon the frequency of sampling. A low sampling frequency increases the likelihood that a server will be found to be providing primary services. The higher the sampling frequency, the more accurate the findings, but the sampling will start to have an impact on the metric itself, because it causes the server to spend more time performing a tertiary service. (The measurement of primary services is, in and of itself, a tertiary service.)

Experientially, a one-minute sampling frequency has been shown to provide a good approximation of primary services activity (compared with subjective observation) without causing a performance impact to the systems being monitored.

IX. Virtual Servers

ScE can be calculated for virtual guests in exactly the same way as for physical servers, although a virtual host is always deemed to be providing primary services unless it has no active guests.

X. Converting ScE to DCcE

For a given data center with a total of m servers, DCcE is calculated by simply averaging the ScE values from all servers during the same time period.

$$DCcE = \frac{\sum_{j=1}^{m} ScE_j}{m}$$

XI. DCcE is Not Comparable Between Data Centers

The secondary and tertiary services factored out before calculating DCcE will vary from data center to data center, and therefore this is not a “one-size-fits-all” metric, and it should not be used to compare different data centers with each other. DCcE is designed to allow server and data center operators to discover where inefficiencies lie within a specific data center and then to address them to increase efficiency over time.

XII. Example Data Set

To prove the veracity of the DCcE metric, TGG used a tool that incorporates the four methods for measuring primary services activity and then used it to determine ScE and DCcE for 350 production servers over a one-month period

As with any sample-based measuring mechanism, the more often samples were taken, the more accurate the result. The tool used in this test collected ScE data every minute and then averaged it across 20 samples. This approach provided sufficient accuracy for tracking primary service activity, did not place undue monitoring overhead onto the systems being monitored, and reduced the data storage requirement to a manageable level.

Data was collected in a central relational database where it could be further aggregated and analyzed.

Figure 1 and Figure 2 illustrate the general lack of correlation found between both overall average CPU utilization and overall average I/O when compared with overall average ScE for each server across the full data set. These findings show that ScE (and therefore DCcE) is not overly influenced by any single underlying primary services measurement.

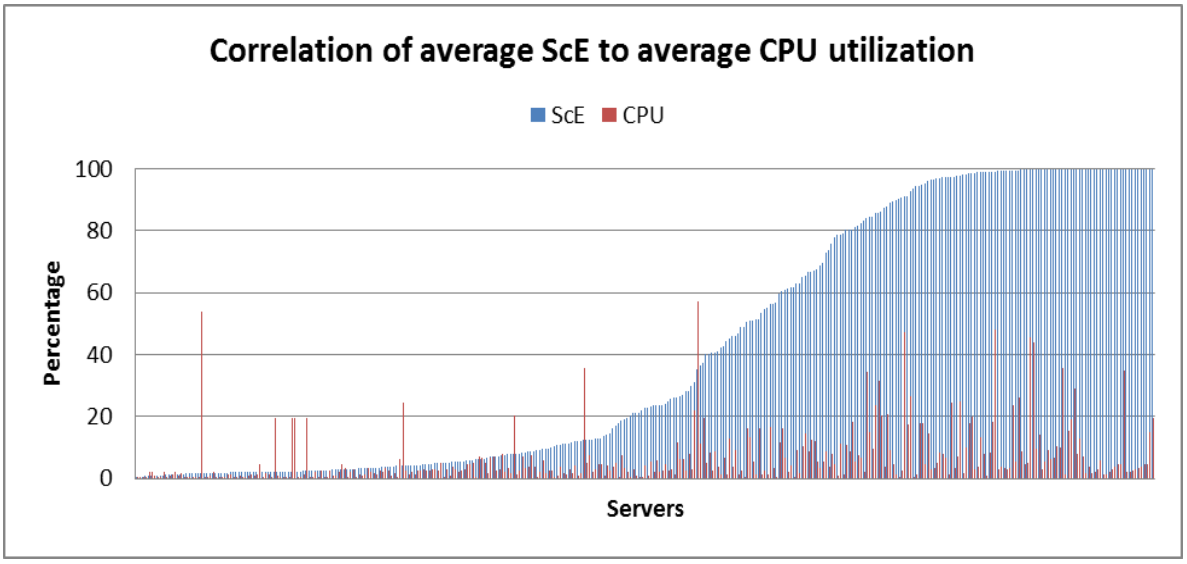

Figure 1. Correlation of average ScE to average CPU utilization

Servers were sorted by their average ScE, and both average ScE and average CPU utilization were plotted. The result reveals that although there is generally a higher CPU utilization for servers with more primary service activity, there are also some servers showing high levels of CPU utilization that had very little primary services activity and some servers with low levels of CPU utilization that were, in fact, continually providing primary services. (See Figure 1.)

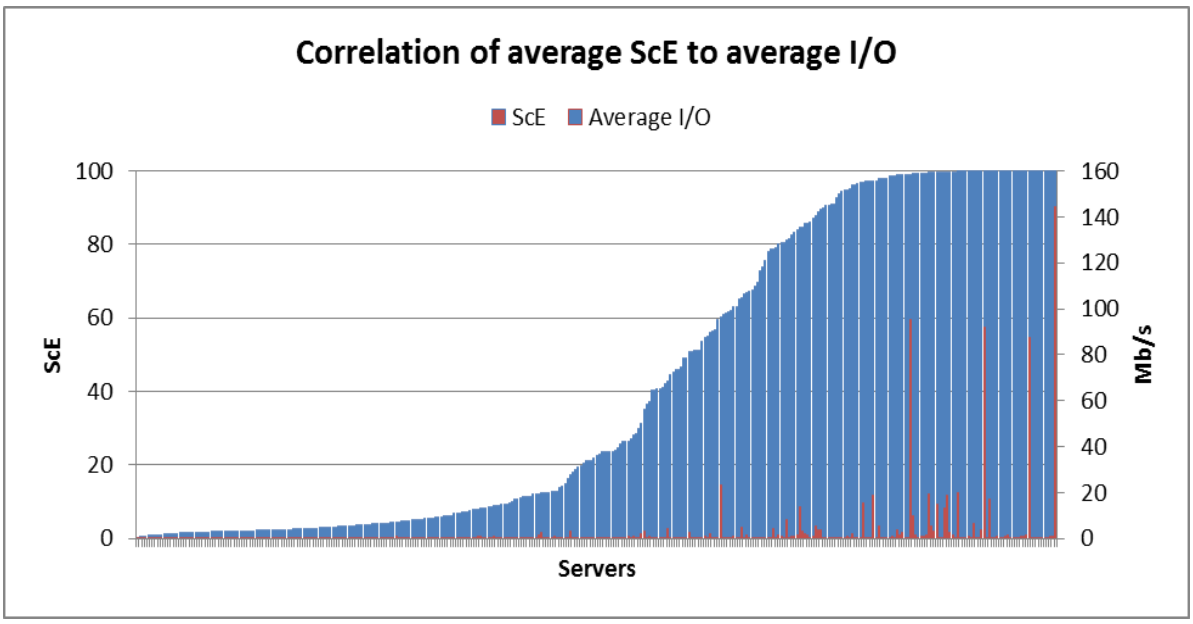

Figure 2. Correlation of average ScE to average I/O

Similarly, there are many servers with a low level of average I/O but high ScE, although no servers were found to be doing large amounts of secondary and tertiary services I/O. (See Figure 2.)

XIII. Issues Discovered During Testing

Since DCcE is comprised of measurements that are easy to automate, tools can be used to capture and analyze the data on a frequent basis. By taking the ScE figure for a particular server and mapping that against measured power data for the same server, it is possible to ascertain the proportion of power being consumed to provide primary services.

The following figures depict a two-week period of time (the tick marks on the x axis represent days), the amount of power being used by a server (the orange areas), and the amount of that power that is being used to provide primary services (the blue areas, found by applying the ScE ratio to the power being used for each particular sample). The green dotted line shows the minimum amount of power used by the server (i.e., when it is idle).

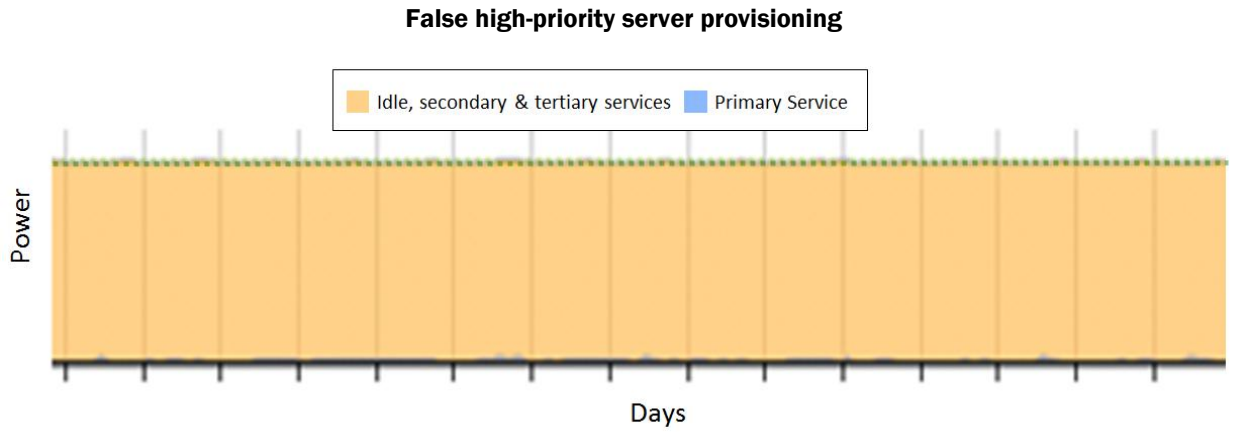

Figure 3. No primary services being provided

Often IT departments are pressured into prioritizing server provisioning tasks, working overtime or neglecting important tasks to silence those who are shouting loudest to get their new server. Several recently provisioned servers in the test group were shown to have no primary services activity at all (the servers used power to idle, not to provide primary services), belying the need to provision them so rapidly. (See Figure 3.)

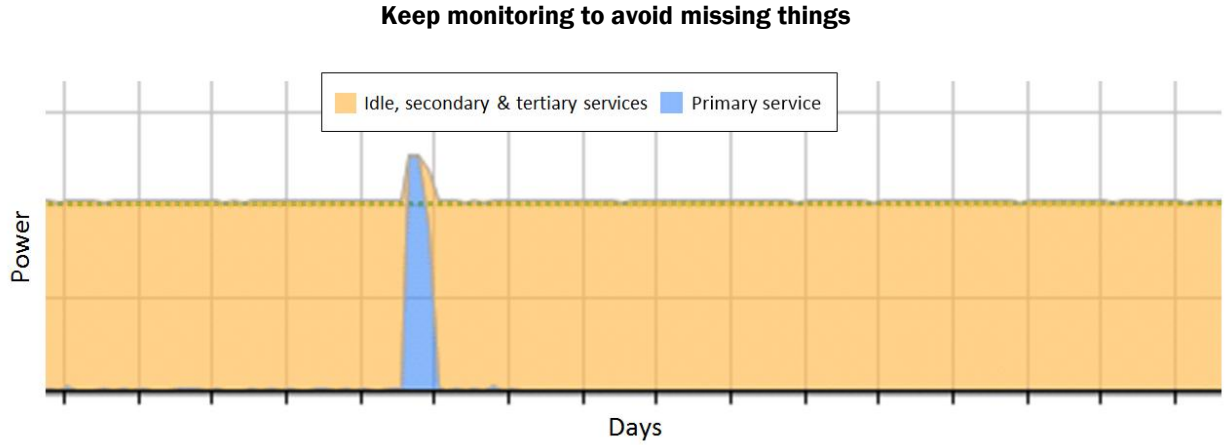

Figure 4. Sporadic primary services activity

It is important to collect sufficient data to satisfy any decision to power down, repurpose, or decommission servers. Figure 4 provides an example of a server that looked as though it may be unused but, upon further monitoring, turned out to be providing sporadic primary services activity.

Collecting two or three months of data enables discovery of servers that undergo periods of infrequent use, such as those that provide important primary services only once a month. Being able to clearly identify where power is being wasted will help data center operators rationalize more aggressive power management, perhaps including the virtualization of servers or the use of change control to power down a server and schedule servers to be powered back on.

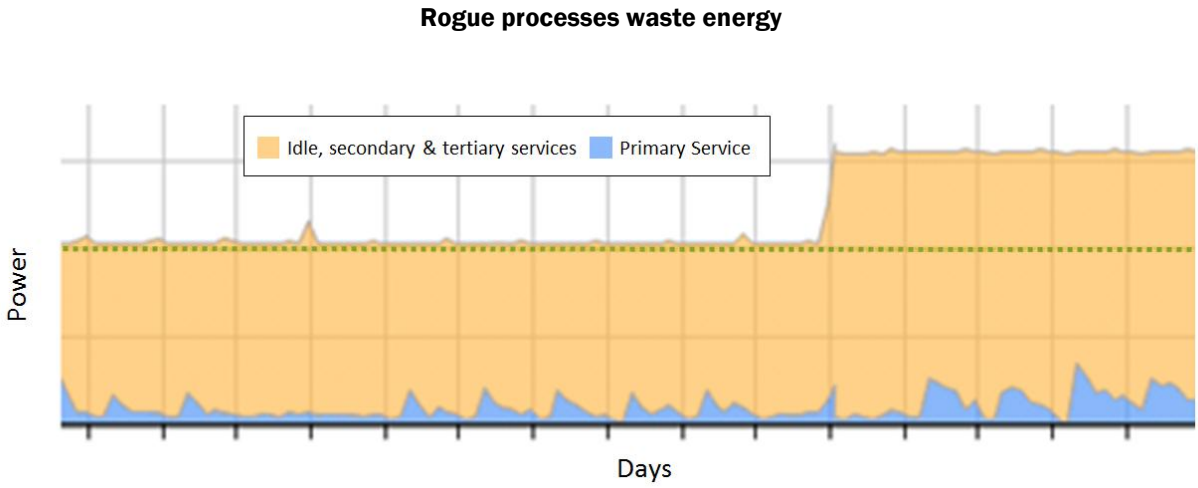

Figure 5. Sudden prolonged changes in behavior

The ScE data revealed sudden and prolonged changes to the power utilization profile of servers. In Figure 5, for example, the amount of primary services activity did not change, and yet the power consumption increased by approximately 100 watts after a point in time.

The relatively low amount of blue (primary services activity) against the y axis (power) reveals that this is an application that does not stress the server. Many servers are over-specified, and because the application's additional load was minimal enough to avoid notice, the primary services performance remained unchanged.

Obtaining further detailed information about the processes running on servers showing this type of behavior can identify which process caused the increase in power consumption. Examples of such processes found during this test include a script stuck in a loop, a device driver with a bug, and a systems management agent with a known issue keeping the processor at 100%.

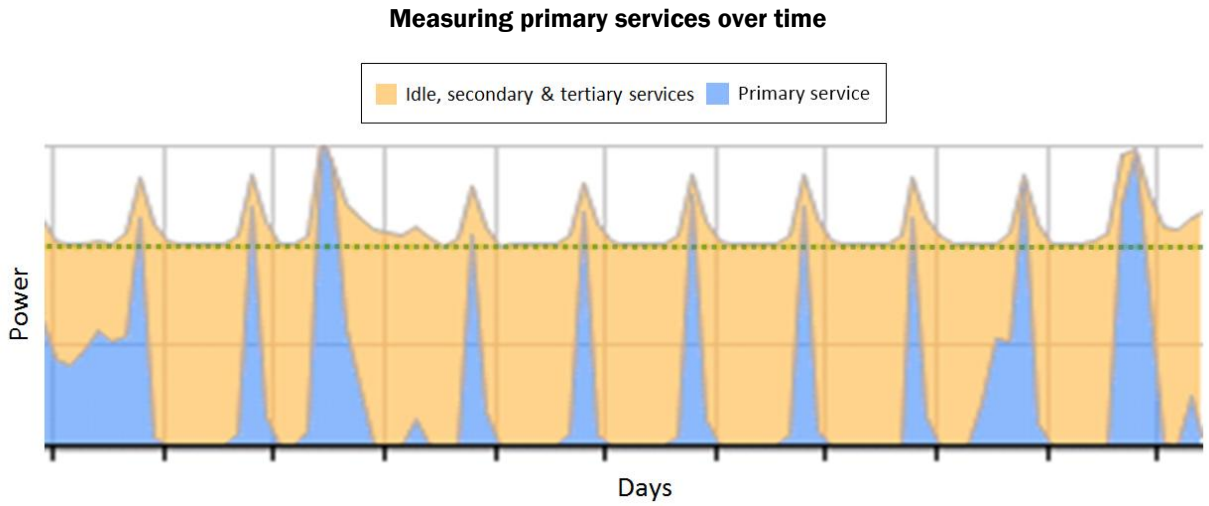

Figure 6. Common production pattern for primary services

Figure 6 illustrates a pattern exhibited by many production servers. It clearly shows regular occurrences of primary services activity (blue) and the periods of time when no primary services are being produced (no blue).

The fact that this example is not quite an exact repeating pattern demonstrates why schedule-based power management for servers does not work—it is not always possible to know in advance when they will be used.

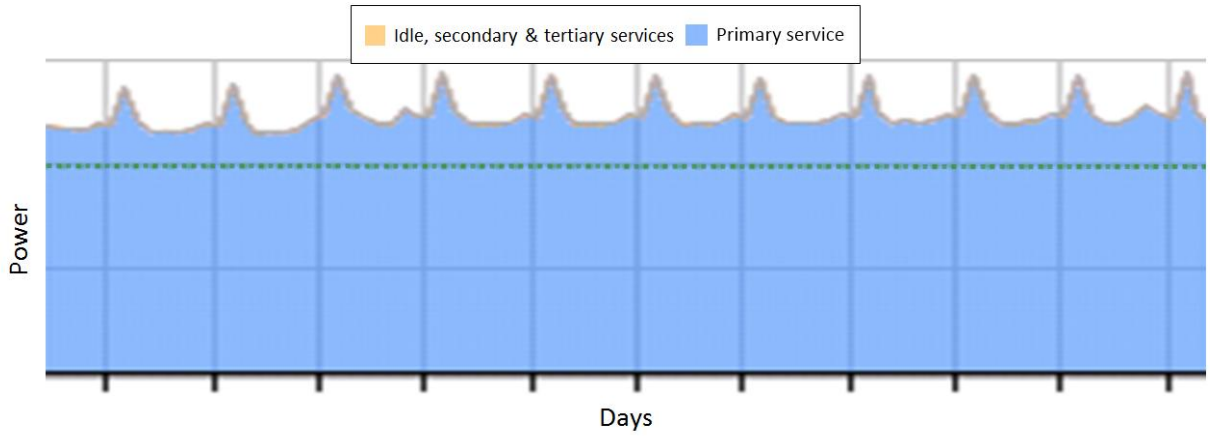

Figure 7. A well-used server

ScE can reveal areas of inefficiency, but it can also show areas of efficiency. Figure 7 is an example of a server that satisfies the measure of primary services against time—it is providing primary services all the time—and against power—all power is devoted to primary services and the server is well-utilized (i.e., power consumption is not at or near idle for that particular server make/model).

XIV. DCcE versus Productivity Proxies

DCcE is not itself a productivity metric; it does not show how much work is being done by the data center.

Some of The Green Grid's proposed productivity proxy metrics already pre-define useful work by only measuring the productivity of known useful business processes. (These include Proxy #1: Useful Work Self-Assessment and Reporting, Proxy #2: DCeP Subset by Productivity Link, and Proxy #3: DCeP Subset by Sample Load.)[2] However, the rest of the proxies only record an amount of work being performed, without any indication of how much of the work is useful.

These productivity proxy types can be used in conjunction with DCcE to provide a measure of how much of the work performed involves primary services. For example, if Proxy #5: Weighted CPU Utilization—SPECint_rate produced a measurement of 10,000 jobs/kWh, and the same set of servers had a DCcE measurement of 50%, then the combined value would be 5,000 primary service jobs/kWh.

XV. Future Oportunities

The Green Grid is in discussion regarding future opportunities around further metrics similar to DCcE for other IT resource types. As an example, The Green Grid may proceed with data center storage efficiency (DCsE), data center network efficiency (DCnE), and other such metrics. Please look to future publications for additional information.

XVI. Summary

Server compute efficiency (ScE) provides a method of calculating efficiency that can be very useful in helping data center managers improve overall energy use. Once managers can see which servers are not providing primary services for prolonged periods (weeks/months), they can switch off, decommission, or virtual those servers. The consequent reduction in demand-side power consumption has a secondary effect on the power consumption of the underlying data center infrastructure.

ScE can be computed using existing, easily obtained performance metrics that are available in all current operating systems. It does not rely on the re-instrumentation of existing applications, making it much simpler to implement than possible alternatives.

Through testing, ScE has been proven to provide an accurate indication of when servers are (or are not) providing primary services, and it also has highlighted areas of concern that would otherwise have gone un-noticed.

By automating the collection of ScE values and averaging them across all servers in a data center, it is possible to derive a metric—data center compute efficiency (DCcE)—that can be used to continually police the overall compute efficiency of the data center. DCcE enables a data center operator to track overall compute efficiency over time and determine whether the data center's server population is right-sized for the job at hand. This consequently makes it possible to ensure that the power and cooling infrastructure is correctly sized to support the necessary load, which in turn allows for optimization of power usage effectiveness (PUE).

XVII. About The Green Grid

The Green Grid is a global consortium of companies, government agencies, and educational institutions dedicated to advancing energy efficiency in data centers. The Green Grid does not endorse vendor-specific products or solutions, and instead seeks to provide industry-wide recommendations on best practices, metrics, and technologies that will improve overall data center energy efficiencies. Membership is open to organizations interested in data center operational efficiency at the Contributor, General, or Associate member level. Additional information is available at www.thegreengrid.org.

---

[1] Unused Servers Survey Results Analysis – The Green Grid White Paper #28
http://www.thegreengrid.org/en/Global/Content/white-papers/UnusedServersSurveyResultsAnalysis
[2] Proxy Proposals for Measuring Data Center Productivity – The Green Grid White Paper #17
http://www.thegreengrid.org/en/Global/Content/white-papers/Proxy-Proposals-for-Measuring-Data-Center-Efficiency

