October 21, 2009

TO: Evan Haines
ICF International
1725 Eye Street, NW, Suite 1000
Washington, D.C. 20006
ehaines@icfi.com

CC: Andrew Fanara
EPA
fanara.andrew@epa.gov

FROM: Chris Hankin
Sun Microsystems, Inc.
chris.hankin@sun.com
202-326-7522

Re: Comments by Sun Microsystems, Inc. on the Energy Star Tier 2 Preliminary Specification for Computer Servers

Dear Mr Haines:

Thank you for the opportunity to provide comments on the EPA's Preliminary Draft for Tier 2 of the Energy Star specification for Computer Servers. Sun Microsystems appreciates the opportunity extended for inclusion in this process, and we look forward to continuing to help achieve a successful new tier of the specification. We especially appreciate your accommodation of our request for more time to complete this response.

We look forward to discussing these points in more detail and to concluding the Tier 2 Energy Star for Servers specification.

Sincerely,

Chris Hankin
Sun Microsystems, Inc.
chris.hankin@sun.com
202-326-7522

Table of Contents

1.Introduction....3

2.Expansion of scope....3

3.Definitions....3

4.Blades....4

5.Net Power Loss....4

6.Active Mode Power Efficiency....4

7.Processor Utilization....5

COMMENTS BY SUN MICROSYSTEMS, INC. ON THE EPA ENERGY STAR (PRELIMINARY DRAFT VERSION 1.0 TIER 2) PROGRAM REQUIREMENTS FOR COMPUTER SERVERS

1. Introduction

Sun commends the EPA on issuing the Preliminary Draft of Tier 2 of the Energy Star for Servers specification. Sun applauds the open process that the EPA is following, including the extensive dialog with the industry and the EPA's willingness to be available for detailed discussions. Sun appreciates the opportunity to meet with the EPA in one-on-one meetings and in industry conference calls, as well as EPA's outreach to the industry at various conferences and symposia.

Sun particularly commends the EPA's transparency throughout the process.

Sun has comments about the following aspects of the Preliminary Draft specification, which are detailed in this document in the following sections:

2. Expansion of scope
3. Definitions
4. Blades
5. Net Power Loss
6. Active Mode Power Efficiency
7. Processor Utilization

2. Expansion of scope

We recommend not to expand the scope of the Server Energy Star Specification beyond 4 sockets, as we do not see sufficient market share for such servers to justify doing this.

3. Definitions

1. In the definition of High Performance Computing System, what does “a large scale construct” mean? It is not clear from this definition how one would differentiate between a large general purpose system with lots of memory and an HPC system.

2. The definition of Server Utilization is very vague. What are the units of ”processor activity”? What does “maximum ability” mean? The Green Grid white paper quoted doesn't give any more details. This is important because looking inside a modern multi-threading processor is rather like looking into reality at the level of quantum physics. Things get nonlinear at that level, where multiple processors share some of the same hardware components, and some of

those components turn themselves off and on based on utilization at the sub-instruction level, borrow resources from each other, etc. Under these conditions counting clock frequency doesn't really account for the more complicated reality.

3. The existing server definition needs to be reconsidered, since, as we have seen from Tier 1, it results in comparing desktop-derived one and two socket low core count servers with high core count enterprise servers. Perhaps we need to subdivide the servers into classes based on other attributes, such as number of cores per socket (<4) or number of DIMMs (<8) on the motherboard.

4. Blades

In response to the discussion of Blades on page 12, the notion of a "fully populated chassis" is a very indistinct one. Since the blade chassis presents a standardized interface, a number of different blade types may be inserted into the chassis, including those containing communications interfaces, disks, etc., as well as blades containing different types of processors or blades with different numbers of processors on them. For instance, Sun Microsystems sells both SPARC and X64 blades. All of these may coexist in the same chassis.

Given this, it is problematic to measure a "fully populated chassis". It would make more sense to measure an empty chassis and a chassis plus one blade installed for each supported type of blade to give the chassis and blade (by subtraction) values.

We do not believe that the blades and other form factors such as rack mount should compete against each other for Energy Star qualification, but that the Energy Star Specification should allow for comparison among different blade products, as there are other factors (such as available space, total capacity needed, ease of reconfiguration or repair, lower cost of redundancy, lower cost of upgrade, etc.) that are likely to drive the decision between buying a rack mount server and a blade server. The customer should be able to evaluate the power efficiency of their proposed purchases in the context of their needs.

5. Net Power Loss

The industry, and its customers, have a long and well developed understanding of power supply efficiency and its measurement. Manufacturers do not add extra power supplies to a system on a whim, but for valid reasons of customer required redundancy. We oppose the introduction of a more complicated measurement that will require customer education and additional testing and reporting.

6. Active Mode Power Efficiency

We continue to believe that an idle power requirement is inappropriate for server Energy Star qualification, and that the increasing cost and limitations of power, the uptake of virtualization technologies, and cultural factors conspire to reduce the relevance of idle power. We also note that the current scheme places a single socket single core print server in the same bucket as a 16 core single socket enterprise server for idle, which does not provide the purchaser with useful information.

We do support a requirement for reporting idle power consumption.

While there are no doubt customers who require general purpose servers, there are also many large customers who have specific performance requirements for specific types of workloads. The latter will not be well served by a measurement of performance on a mixed bag of workloads.

We suggest that it should be possible for a server or family of servers to be qualified based on superior power efficiency on a specific type of workload. To this end we recommend that, in the language of the discussion guide, a combination of approaches A & C be adopted. Or, another way of saying this, is to adopt Approach A, but allow the "Specialized Efficiency Rating Tool" to be one of the benchmarks that may be used to qualify a server under Tier 2.

This allows customers to find qualified products that meet their needs for specific workload types or, if they don't have specific workloads, to find a qualified general purpose server.

As an alternative to forcing all server manufacturers to publish benchmark results for which their servers may not do well, a qualifying score could be some percentage of the maximum published benchmark power efficiency score. To make this work, it would probably have to be the case that qualification based on a specific power efficiency score for a benchmark would have to last for some minimum time even if the bar is raised by a subsequently published high score, to provide some stability in the process.

We also recommend that the efficiency rating tool should include a virtualization component, since virtualization is likely to be the best way to achieve server consolidation - which will give the best overall power savings.

7. Processor Utilization

The EPA's requirement for processor utilization measurements of up to 5% accuracy, using a particular algorithm, is extremely difficult to meet, and unnecessary for the following reasons:

1. Each processor and operating system that ships today already has built-in techniques to calculate processor utilization. These techniques and algorithms differ from processor to processor and from OS to OS, but they all yield reasonably accurate data for the purpose of decision making at the data center level. The imposition of a particular government specified formula for calculating processor utilization remains highly problematic, as it precludes any changes to the shipping systems. The imposition of a particular algorithm for calculating processor utilization also stifles innovation and improvement in measurement techniques.

2. There is continued innovation in power management at the microprocessor level which includes technologies like dynamic voltage scaling, dynamic frequency scaling, core power reduction, core disabling, cycle skipping, slower clocking, halt states, and several others. The algorithms for measuring processor utilization will continue to undergo ongoing innovation and refinement to account for these new and upcoming technologies. Any option to provide a defined algorithm to improve reporting accuracy will never be able to account for the variability and range in processor power management techniques.

3. The intent of making processor or system utilization available to the data center operators is to encourage them to track the use of their equipment and identify equipment that is not utilized or under-utilized. To enable this, the measurement only needs to be sufficiently accurate for the purpose of enabling decisions around the reprovisioning of workloads or the migration of virtual machines. Existing CPU utilization measurement algorithms are already sufficiently accurate for this purpose. Many data centers today already rely on CPU utilization numbers as reported by currently shipping operating systems on currently shipping servers. They use this information successfully today to dynamically manage data center power consumption by reprovisioning workloads to minimize the under-utilization of machines. Requiring a specific accuracy criterion for CPU utilization will not provide any particular incremental value to customers, and will only increase the cost of the system due to the expensive additional micro-instrumentation required. In addition, it will be very difficult to tell what accuracy is actually achieved.

Request:

Remove the reporting accuracy requirement for processor utilization. Do not mandate any particular algorithm or formula for calculating processor utilization.