USITO preliminary response to the CNIS Draft Server Energy Efficiency Standard (April 2018):

June 8, 2018

USITO and the TGG SERT Working Group (SERT WG) appreciate the opportunity to review and comment on the CNIS "Minimum Allowable Values of energy efficiency and energy efficiency grades for servers". The current version offers the beginnings of the structure to the CNIS standard for energy efficiency for servers. Based on a review of the document we have the following observations and concerns:

1. USITO encourages CNIS to use the SERT metric test data and overall test score for the assessment of server energy efficiency. The SERT test offers a proven, industry supported metric that designates a specific set of testing requirements that tests all server products equally and consistently in a way that provides an accurate, and fair relative assessment of a server's energy efficiency as measured by workload delivered per unit of energy consumed.
2. CNIS has proposed an alternative metric using worklet efficiency values calculated by taking the mean of the quantity interval values of the measured power times elapsed test time divided by the un-normalized transaction count. The alternate calculation may present some complications to assessing server energy efficiency. USITO is proposing a data analysis process, on which the SERT WG has already started work, which will assess the effectiveness of the proposed CNIS metric. USITO proposes to provide this document to CNIS by July 13, 2018.
3. The comments below provide specific suggestions additions and edits to the draft standard.

The CNIS choice to use energy per transaction may provide a workable approach to setting energy efficiency requirements for servers. Based on our initial assessment the metric should be roughly equivalent to the inverse of the SERT score where the raw transaction or performance data is normalized for a reference system. However, we are concerned that introducing elapsed time, reducing the number of worklets, not normalizing the scores using a reference server, and inverting the performance to power ratio used in the SERT metric may change the relative assessment and ranking of the different servers. Another concern is that assessing server efficiency differently from SERT methodology will impose additional conformity assessment burden on server manufactures and the labs, including the ones in China, when shipping server products globally.

Given this approach, USITO is providing our preliminary observations/concerns in regards to the April 2018 draft. In addition, USITO with the SERT WG have initiated additional analysis activities with regards to the metric proposed by CNIS; these activities are also described in this document.

Observations and Concerns

USITO has carefully reviewed the CNIS draft server energy efficiency regulation. Based on our review of the regulation we recommend the following edits and additions to the regulation to improve its clarity and implementation.

Section 3.2, 4.2 and the regulation title should refer to the "maximum allowable typical energy consumption value" for servers. Section 5.2 defines the worklet efficiency values as energy consumption per transaction which would be calculated as the average of the four interval values calculated by combining the measured power times the elapsed time divided by the transaction count for each interval. If our understanding of the metric is correct, a lower energy consumption per transaction will represent a higher efficiency. Thus, grade 3 in Table 1 should be the maximum

allowable energy consumption per transaction and the grade 2 and grade 1 thresholds should be lower than the grade 3 threshold.

Mathematical Relationship between SEE and SERT scores: Combining the elapsed time with the power measurement and dividing by the transaction count creates the inverse[1] of the SERT CPU and storage worklet efficiencies. The proposed CNIS metric has units of power-time/transaction count while the SERT metric has transaction count divided by power-time. Differences in the relative values will also be introduced if CNIS were to use the raw transaction count rather than the transaction count normalized to a reference system like in SERT.

The one worklet which is different in SERT compared to the CNIS April draft is the Flood memory score. For the Flood memory score the actual workload measurement which affects power demand is the score, which measures the bandwidth used to complete the test. The transaction count is set to match the number of transactions to the number of virtual Logical Partitions (LPARs), which is what is required to drive use of the full memory bandwidth in the server. The transaction counts are not indicative of the power use. To account for this difference, $SEE_{Mem50/100,}$ CNIS equations 11 and 12, may best be calculated as described in Equation 1. However, the SERT WG is still working on the best calculation method for the memory worklet.

$$SEE_{Mem50\%\ or\ 100\%} = \frac{Power\quad (W) * Elapsed\ Measurement\ Time\ (s)}{Flood\ Score}$$

Where:

$SEE_{Mem50\%\ or\ 100\%}$ = Memory interval efficiency score

Power = Average Power of server measured at the interval

Elapsed Measurement Time = SERT reported elapsed in seconds

Flood Score: The Flood performance score as reported in the SERT results_details.html datasheet.

Equation 1: Calculation Method for SEE_{Flood} interval efficiency using SERT measurements.

For the purposes of the CNIS SEE_{MemXX} and E_{Server} values and because the unique nature of the Flood worklet as compared to the CPU and storage worklet scores, USITO and the SERT WG will investigate how best to apply the SERT Flood memory worklet measurements to create a worklet score that matches the intent of the CNIS approach. This may result in an adjustment to the equation provided above when we make our July 13, 2018 report.

Configurations for testing: The draft standard is silent on the configurations to be tested to demonstrate conformance to a specific server efficiency grade. Clearly defining the tested configurations is essential to being able to accurately compare the efficiency of two or more different servers due to the range of efficiency scores possible as the server configuration is changed. USITO recommends that a company test three server configurations – Low-end performance, High-end

[1] This assumes the equivalency is assessed using scores normalized with a reference system.

performance and Typical - to demonstrate conformance to the energy efficiency grades proposed by CNIS. The three configurations would be defined as follows:

"High-end" performance configuration: The "high-end" configuration is comprised of a minimum of two solid state drives (SSDs), with a minimum of 3 times the Calculated Memory Capacity of the EUT and with the Calculated Processor Capacity which represents the highest performance product model within the server product family.

"Low-end" performance configuration: The "low-end" configuration is comprised of a minimum of two 10.000 rpm 3,5" HDDs, with a minimum of 1 times the Calculated Memory Capacity of the EUT and with the Calculated Processor Capacity which represents the lowest performance product model within the server product family.

"Typical" configuration: A product configuration that lies between the Low-end Performance and High-end performance configurations and is representative of a deployed product with a high volume of sales. The configuration will have at least two times the Calculated Memory Capacity.

All memory channels must be populated with the same DIMM raw card design and capacity per the requirements of the SERT test.

The Processor Capacity and Minimum Memory Capacity are calculated as follows:

EXAMPLE 1: Calculation of Processor Capacity:

Processor Capacity (Dimensionless) = "the number of central processor units (CPUs)" * "the number of cores per CPU" * "the number of threads per core" * "processor frequency"

For a system with 2 CPUs, each CPU has 4 cores and 2 threads per core, frequency of 2.2;

Processor capacity = 2*4*2*2.2 = 35.2

EXAMPLE 2: Calculation of Memory Capacity:

Calculated Memory Capacity (GB) = $\#\ of\ CPUs \times \#of\ memory\ channels\ per\ CPU \times smallest\ DIMM$

For a system with 2 CPUs, 4 memory channels, and the lowest capacity DIMM is 2GB;

Minimum memory capacity = 2x4x2GB_DIMMs = 16 GB

Section 4.1: USITO continues to be concerned with the proposal to set three grades for server efficiency which will result in four product category groupings under the CNIS requirements. USITO has identified the following three issues with the grading approach proposed in Section 4.1.

A single product family will have configurations that get different "Grades": The range in E_{Server} values across the range of server configurations in a single product family (machine type/model) will cause a single server family to have configurations which exist in different grade levels depending on the configuration of the server. A product family could have some configurations which do not meet the grade 3 limit (would not be eligible to be placed on the market) and other configurations could meet grade 1 limit (representing the most efficient servers). The 3 level grading system will make it very complicated to certify servers to the China market.

One example: Initial analysis of a 7 CPU worklet E_{server} calculation on 13 server families shows an average ratio of the active efficiency of the lowest efficiency server to highest efficiency server of approximately 3.4. For one particular product family, the ratio between the active efficiency score for a low-end performance and high-end performance configurations is over 17, meaning that it is likely that this server will have configurations with Eserver scores in each of the three grading categories.

USITO will assess the dataset and propose one or more approaches for managing server product families under a grading system in the July 13, 2018 submittal to CNIS (USITO also plans to assess the additional issues that are described in the USITO Data Analysis Plan, provided below.)

A category for blade/multi-node servers needs to be added: The SERT WG has analyzed the SERT active efficiency scores by server form factor and socket count. That analysis has identified the fact that rack servers and blade servers have materially different SERT active efficiency scores when comparing servers with the same product launch year and configuration type. Table 1 shows the average SERT active efficiency score for rack and blade server products matched by the year the server product was put on the market and the configuration type. In general, the blade server has a 10-25% higher efficiency score relative to the rack server for a given launch year and configuration type. There are three exceptions to this (2014, High-end performance and 2013 and 2017 Low-end performance) in the data set. The differences in the 2013 and 2017 high-end configurations can be explained by the limited number of configurations in the blade server sample. For the 2014 High-end performance difference, it is likely the result of the choice of processor and component types for the blade servers as compared to the rack servers. Because of the material difference in the blade and rack scores, USITO recommends that the grading system have three form factor categories: Tower, Rack and Blade/Multi-node (referred to as "Blade" in Table 1).

			Average SERT Active Efficiency Score by Configuration type by Launch Year for 2 Socket Rack and Blade Servers					
Server Launch Year	Configuration Count		High-end Performance Configuration		Typical Configuration		Low-end Performance Configuration	
	Rack	Blade	Rack	Blade	Rack	Blade	Rack	Blade
2010								
2011		3		5.5		6.4		6.4
2012	2	24	7.8	8.7		8.6	4.9	5.6
2013	21	8	9.1	10.0	8.0	9.0	6.5	6.2
2014	62	31	15.8	13.2	10.4	11.6	9.1	8.6
2015	8	12	15.5	19.1	13.3	15.2	12.4	15.0
2016	7	0	14.7		15.4		17.1	
2017	24	7	19.4	31.4	14.2	33.4	13.4	11.8
Total	124	82						

Table 1: Average SERT Active Efficiency Score by Configuration Type by Product Launch Year for 2 Socket Rack and Blade Servers

There are too many energy efficiency grades for the limited number of server product families: Table 2 shows the number of server product families and configurations in the ITI/TGG database for 1 and 2 socket server products. While the dataset does not include all the servers available on the market, it provides a general indication of the relative number of server products in each form factor and socket count category. The data suggests that there are insufficient product families available to support a 3 grade server energy efficiency standard for 1 and 2 socket tower servers and 1 socket rack and blade servers. Even with the two socket rack and blade systems, there will be a limited number of products in each of the 4 product groupings created by a three grade system. USITO recommends that CNIS consider a two grade system with market entry and voluntary "top runner" thresholds for products.

		1 socket products			2 socket products		
		Tower	Rack	Blade	Tower	Rack	Blade
Number of Product Families		3	4	0	4	41	28
Number of Configurations		10	16	0	16	206	143

Table 2: Number of Product Families and Configurations in the ITI/TGG Database

Sections 5.2 to 5.4: The equations show the worklet efficiency scores being calculated using the arithmetic mean. USITO recommends that CNIS combine the worklet efficiency scores or the power*elapsed time (numerator of the workload term) and transaction counts (denominator of the workload term) using the geometric mean function. The use of the geometric mean function smooths out the impacts of highly varied values, preventing a specific interval or worklet from unduly influencing the final score.

5.3 Calculation of Server energy efficiency when memory is in work mode: The calculation of the memory efficiency calls for the numerator to be the "number of server operations at different memory workloads". For the Flood worklet, the number of transactions is set to match the number of available threads to insure that all (100% interval) or half (50%) interval is exercised by the test. The indicative parameter for the memory test is the test score itself – which is the measured bandwidth in GB/s. The measured bandwidth, not the transaction count, needs to be used as the value in the numerator of the worklet efficiency equation to accurately represent the work being performed in the memory worklet test.

USITO Data Analysis Plan

Based on the USITO assessment of the proposed SEE metric, USITO has identified two specific areas of concern with how the metric is constructed and calculated: (1) the use of only 2 CPU worklets may not appropriately assess the work capacity and capability of a server and (2) the use of the un-normalized transaction counts (transaction counts have not been normalized to a reference system as was done in the SERT test) may allow one or two worklets to be overly influential in the calculation of the overall SEE score. In order to assess the importance and impact of these items on the assessment of server energy efficiency, USITO will undertake a set of data analysis projects, which are detailed below, to assist CNIS in evaluating the effectiveness of the SEE metric.

Separately, it is also important that CNIS designate the elapsed time, not the transaction time, to calculate the server energy consumption value for the worklet tests.

Calculation of the server energy value: In order to calculate energy use, it is important to use the elapsed time (column 5 in table 3) and not transaction time (column 8 in table 3). The transaction time is a value of the cumulative time that all of the threads are run during the SERT interval (thread-time). It is not representative of the time that the test is run on the server. Table 3 shows the results-details report for the Compress worklet from one configuration.

Phase	Interval	Actual Load	Score	Elapsed Measurement Time (s)	Transaction	Transaction Count	Transaction Time (s)		
Warmup	max		23,252.50	30.289	TxCompress	703,416	1,341.00		
Calibration	max		24,076.86	120.29	TxCompress	2,896,182	5,537.95		**Proposed CNIS metric**
	max		24,051.75	120.312	TxCompress	2,893,452	5,539.89		
	Calibration Result		24,063.15					Worklet power demand (watts)	energy / transaction counts (w-s/ trans)
Measurement	100%	99.70%	23,979.92	120.289	TxCompress	2,884,515	5,469.09	936.9	0.039070
	75%	75.00%	18,051.22	120.248	TxCompress	2,170,521	3,663.49	885.8	0.049074
	50%	49.90%	12,010.38	120.228	TxCompress	1,443,973	2,440.03	805.8	0.067092
	25%	25.00%	6,023.72	120.242	TxCompress	724,299	1,235.38	721.3	0.119744
						Geometric mean of Compress worklet scores			**0.062648**

Table 3: Example of Compress measurement data and Proposed CNIS workload score calculation

Availability of Transaction Count and Elapsed Time data: A complication for the SERT Working Group is that we do not have transaction counts or elapsed time in our dataset. Those two data points are not reported in the .xml file that we have used to populate the dataset. In order to do analysis with the current dataset, we are going to need to hand load the transaction count and the elapsed time data. We plan to do analysis on the 2 socket rack servers for server product released in 2016 and 2017 where we can get the results.details data sheets from the manufacturers.

If CNIS intends to use the transaction counts and elapsed time data to calculate the SEE metric, USITO recommends that CNIS approach SPEC to include those data points in the .xml file. This will simplify the data extraction from the SERT test results.

Assessment of the impact of using 2 CPU worklets on server ranking and un-normalized transaction count or performance scores to calculate the SEE metric:

We propose calculating the following server efficiency scores using the ITI/TGG SERT data set for each of the server configurations for products released in 2016 and 2017 where we have access to the transaction count data.

a. An E_{server} value using the un-normalized SERT Compress and LU, Flood and Random read/write worklet scores. This is the metric proposed in the CNIS draft standard.

b. An E_{server} value using the normalized SERT Compress and LU, Flood and Random read/write worklet scores.

c. An E_{server} value using the un-normalized SERT Compress, LU, CryptoAES and Hybrid ssj, Flood and Random read/write worklet scores.

d. An E_{server} value using the normalized SERT Compress, LU, Flood and Random read/write worklet scores.

e. An E_{server} value using the normalized SERT CPU worklets (7), Flood and Random read/write worklet scores.

f. SERT active efficiency metric.

Using the E_{server} values calculated above and the SERT active efficiency score, we will compare the relative ranking of servers by calculated efficiency value and deployed power value in the data set as follows:

a. The SERT active efficiency score and the proposed E_{server} value for the two CPU and 4 CPU workload scores for both normalized and un-normalized scores.

    i. As part of this analysis, the changes in the three workload scores – CPU, memory and storage - will be assessed to determine which workload changes are causing changes in the overall ranking score, if any.

b. Compare the relative results of the metrics for the two CPU and four CPU workload scores calculated using normalized and un-normalized transaction counts.

c. Compare the relative ranking of the results for the 2, 4 and 7 CPU workload scores calculated using normalized transaction counts to ascertain what benefits are achieved by using more CPU worklets.

d. Performing these comparisons will enable the SERT WG to assess the impact of using the CNIS calculation method, different numbers of CPU worklets and normalized and un-normalized scores on ranking changes and server qualification in China.

Sensitivity Analysis of E_{server} score to the number of CPU worklets: Consistent with USITO's previous analyses and comments to CNIS, USITO continues to be concerned with the use of only two CPU worklets to create the CPU score. Using only two worklets increases the potential that a manufacturer will create an accelerator for one of the two workloads to substantially increase the overall score based on the acceleration of a single worklet. As discussed in previous communications with CNIS, the SERT WG recommends that at least 4 CPU workloads be used to minimize the potential for the acceleration of a single CPU worklet to overly influence the overall SERT score.

To demonstrate this risk, we will use data from SERT tests on 2012-2013 products where some servers had enabled the crypto-acceleration capability in their jvms while others had not. This will enable us to assess the E_{server} values for servers with and without the Crypto accelerator to show the relative impact of the accelerated CryptoAES score on a two, four and seven CPU worklet E_{server} score.

Annex B: Standard workloads:

The standard workload descriptions offered in Annex B are sufficiently broad to cover the SERT test and other tests which may be available. However, there is a concern that the description is so broad that it would allow a server manufacturer to write their own code in accordance with Annex B requirements in a way that creates a maximized and optimized score for their set of products. None of these operations are that complicated, so it wouldn't be a difficult problem.

As one example, the memory workload (i.e. Flood) documents each operation, but doesn't describe the specific data for operation of the test. This lack of operating details means that a manufacturer could allocate small arrays that fit in L2 cache and do all of the computations on those cached arrays and hardly touch main memory.

This is one of the biggest reasons that SPEC provides the implementation details that are found in SERT, and not just a general specification. While the SERT code cannot be optimized perfectly for all possible customer applications, the fact that it exercises a range of distinct workloads requires server manufacturers to provide a hardware/firmware/software system that delivers both high performance and energy efficient computing. Overall, SERT provides a fair assessment of the ability to a server to deliver more workload/computational capacity for each unit of energy consumed.

That is why USITO supports the use of SERT and the implementation of the ISO standard for Server Energy Efficiency Metric (SEEM): it designates a specific set of testing requirements that tests all server products equally and consistently in a way that provides an accurate, relative assessment of a server's energy efficiency as measured by workload delivered per unit of energy consumed.

Conclusion:

Given the fact that the SERT test is well established and proven, it is demonstrated to provide a fair and consistent assessment of server energy efficiency attributes, and that companies will have to generate SERT test data for other regulatory and procurement programs, USITO continues to encourage CNIS to designate the SERT test as the test metric for the CNIS "Minimum Allowable Values of energy efficiency and energy efficiency grades for servers" requirements.

While USITO would prefer that CNIS use the SERT test metric active efficiency value for the energy efficiency grades for servers, USITO is partnering with the Green Grid SERT working group to evaluate the alternative calculation methods proposed by CNIS in the April 2018 draft requirements document. The work that we have detailed above will assess how the CNIS proposed value, and some alternative variations on the calculation method, compare to the ranking of the SERT active efficiency metric and whether it results in lower deployed power in the data center for systems chosen as most efficient. USITO will provide CNIS the assessment document by July 13, 2018.

We welcome any comments or questions from CNIS on the information in this document.