Draft ETSI EN 303 470 V0.0.6 (2017-06)

Environmental Engineering (EE);
Measurement Process for Energy Efficiency KPI
for Servers

STABLE DRAFT

Reference

DEN/EE-EEPS24

Keywords

broadband, energy management, ICT, sustainability

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

List of figures

List of tables

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://ipr.etsi.org).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This European Standard (EN) has been produced by ETSI Technical Committee Environmental Engineering (EE).

Proposed national transposition dates	
Date of latest announcement of this EN (doa):	3 months after ETSI publication
Date of latest publication of new National Standard or endorsement of this EN (dop/e):	6 months after doa
Date of withdrawal of any conflicting National Standard (dow):	6 months after doa

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

“must” and “must not” are NOT allowed in ETSI deliverables except when used in direct citation.

Introduction

The present document specifies a metric for the assessment of energy efficiency of computer servers using reliable, accurate and reproducible measurement methods, which take into account the recognised state of the art.

The specification formalizes the metrics representing a single figure of merit of a single computer server representing the relative efficiency and power impact at an ICT site level of deployment. The metric is targeted for use as a tool in the selection process of servers to be provisioned for an ICT site.

The specification formalizes the tools and conditions used to determine a figure of merit for a single computer server representing the relative efficiency and the power impact at a ICT site level of deployment. For comparisons, evaluations should be conducted across similar server types or categories. The efficiency metric is targeted for use in a pass/fail selection process of servers to be provisioned for general purpose operations in a ICT site. The specification does not prescribe the levels or values for acceptance but prescribes standard method of evaluation from which energy efficiency programs would use to establish such criteria.

As there are specialized ICT sites and specialized equipment and configurations, a metric that evaluates provisioning impacts to general purpose operations may not be applicable. ICT site equipment and servers in particular, are generally customized and commissioned on site for deployment. As with most IT equipment, new technologies are regularly introduced, which may require product level customization or an industry wide tool upgrade to more appropriately represent the efficiency of the servers. This specification defines classes of servers to address applicability, configuration groupings to represent a family of servers to address custom configurations, and tool revision control to ensure comparability and consistency of the resulting metric value.

The present document is based upon the Server Efficiency Rating Tool™ (SERT™) of the Standard Performance Evaluation Corporation (SPEC) and takes into account:

- the Ecodesign Technical Assistance Study on Standards for Lot 9 Enterprise Servers and Enterprise Data Storage;
- energy efficiency KPIs standardisation in ISO/IEC JTC/1 SC/39 and CLC TC 215;
- activity related to deployed power by The Green Grid;
- Energy Star for Computer Servers.

This work item defines energy efficiency metrics and measurement methodology for server equipment under standardisation Mandate M/462 of the European Commission. The document does not address home servers and small servers under the Mandate M/545.

1 Scope

The present document specifies a metric, test conditions, and product family configuration for the assessment of energy efficiency of computer servers using reliable, accurate and reproducible measurement methods. The metric applies to general purpose 1 and 2 socket computer servers, with their own dedicated power supply, provisioned for use within a ICT site of which the data processing requirement exceeds the use of a single server.

The metric applies to a computer server product family, including type and count of CPU, memory, storage, power supplies, cooling (e.g. fans) and any other add-on hardware expected to be present when deployed.

The metric applies to the following types of rack-mounted or pedestal form-factor servers:

- blade server with no more than four processor sockets per blade;
- multi-node with no more than four processor sockets.

The present document defines:

- an energy efficiency metric to support procurement requirements;
- requirements for equipment to perform the measurements and analysis;
- requirements for the measurement process;
- requirements for the management of the KPI calculation;
- operation or run rules to configure, execute, and monitor the testing;
- documentation and reporting requirements;
- a validation process for the metric to ICT site level impact.

The present document is not applicable to:

- fully fault tolerant servers;
- High Performance Computing (HPC) systems;
- hyper-converged servers;
- large scale servers;
- multi-node servers;
- network equipment;
- network servers;
- server appliances;
- storage products including blade storage;
- storage servers.

NOTE: Products whose feature set and intended operation are not addressed by active mode testing parameters are excluded from this evaluation method. The above list shows products for which SERT™ efficiency evaluations are not appropriate.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are necessary for the application of the present document.

[1] CENELEC EN 62623:2013, Desktop and notebook computers. Measurement of energy consumption

[2] ETSI EN 300 132-3-1, Environmental Engineering (EE); Power supply interface at the input to telecommunications and datacom (ICT) equipment; Part 3: Operated by rectified current source, alternating current source or direct current source up to 400 V; Sub-part 1: Direct current source up to 400 V

[3] SERT™ Design Document V.2
https://www.spec.org/sert2/SERT-designdocument.pdf

[4] SERT™ User Guide V.2
https://www.spec.org/sert2/SERT-userguide.pdf

2.2 Informative references

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] CENELEC EN 60297 series, Mechanical structures for electrical and electronic equipment. Dimensions of mechanical structures of the 482,6 mm (19 in) series

[i.2] EPRI/ECOVA, Generalized Test Protocol for Calculating the Energy Efficiency of Internal A.c.-D.c. and D.c-D.c Power Supplies Revision 6.7, March 2014 .

[i.3] IEEE 1100, IEEE Recommended Practice for Powering and Grounding Electronic Equipment

3 Definitions, symbols and abbreviations

3.1 Definitions

For the purposes of the present document, the following terms and definitions apply:

active state: operational state of a computer server (as opposed to the idle state) in which the computer server is carrying out work in response to prior or concurrent external requests (e.g. instruction over the network) and includes both active processing and data seeking/retrieval from memory, cache, or internal/external storage while awaiting further input over the network

auxiliary processing accelerator: computing expansion performance enhancement feature installed in general-purpose computing systems

NOTE: an auxiliary processing accelerator is an enhancements to the system, is attached to I/O ports and is not a stand-alone CPU in the system

blade chassis: enclosure that contains shared resources for the operation of blade servers, blade storage, and other blade form-factor devices. Shared resources provided by a chassis may include power supplies, data storage, and hardware for d.c power distribution, thermal management, system management, and network services

blade server: computer server, designed for use in a blade chassis, that is a high-density device and functions as an independent computer server and includes at least one processor and system memory, but is dependent upon shared blade chassis resources (e.g. power supplies, cooling) for operation

NOTE: A processor or memory module that is intended to scale up a standalone server is not considered a blade server.

multi-bay blade server: blade server requiring more than one bay for installation in a blade chassis

single-wide blade server: blade server requiring the width of a standard blade server bay

double-wide blade server: blade server requiring twice the width of a standard blade server bay

half-height blade server: blade server requiring one half the height of a standard blade server bay

quarter-height blade server: blade server requiring one quarter the height of a standard server bay

multi-node blade server: blade server which has multiple nodes

NOTE: The blade server itself is hot swappable, but the individual nodes are not.

blade storage: storage device that is designed for use in a blade chassis and that is dependent upon shared blade chassis resources (e.g. power supplies, cooling) for operation

blade system: system comprised of a blade chassis and one or more removable blade servers and/or other units (e.g. blade storage, blade network equipment) which provide a scalable means for combining multiple blade server or storage units in a single enclosure, and are designed to allow service technicians to easily add or replace (hot-swap) blades in the field

buffered double data rate (DDR) channel: channel or memory port connecting a memory controller to a defined number of memory devices (e.g. dual in-line memory modules (DIMMs)) in a computer server

NOTE 1: A typical computer server may contain multiple memory controllers, which may in turn support one or more buffered DDR channels.

NOTE 2: Each buffered DDR channel serves only a fraction of the total addressable memory space in a computer server.

computer server: computer that provides services and manages networked resources for client devices (e.g. desktop computers, notebook computers, thin clients, wireless devices, Personal Digital Assistants, IP telephones, other computer servers, or other network devices)

NOTE 1: A computer server is sold through enterprise channels for use in ICT sites and office/corporate environments.

NOTE 2: A computer server is primarily accessed via network connections, versus directly-connected user input devices such as a keyboard or mouse.

direct current server: computer server that is designed solely to operate on a direct current (d.c.) power source

hyper-converged server: highly integrated enterprise device which contains the same components as a computer server in addition to the features of a network server and storage server

fully fault tolerant server: computer server that is designed with complete hardware redundancy, in which every computing component is replicated between two nodes running identical and concurrent workloads (i.e. if one node fails or needs repair, the second node can run the workload alone to avoid downtime) and that uses two systems to simultaneously and repetitively run a single workload for continuous availability in a mission critical application

large scale server: resilient/scalable server which ships as a pre-integrated/pre-tested system housed in one or more full frames or racks and that includes a high connectivity input/output subsystem with a minimum of 32 dedicated input/output slots

managed server: computer server that is designed for a high level of availability in a highly managed environment

NOTE: A managed server typically contains an installed dedicated management controller (e.g. service processor) and redundant power supplies.

multi-node server: computer server that is designed with two or more independent server nodes that share a single enclosure and one or more power supplies and in which power is distributed to all nodes through shared power supplies

NOTE: Server nodes in a multi-node server are not designed to be hot-swappable.

network server: large network device which contains the same components as a computer server together with more than 11 ports, has a total line rate throughput of greater than or equal to 12Gb/s and is designed to dynamically reconfigure ports and speed and to support a virtualized network environment, software defined networking

NOTE: Supporting features are described by the product's datasheet description and is either accompanied with vendor specific utilities and/or commercially available software supporting these functions.

pedestal server: self-contained computer server that is designed with power supply units, cooling, input/output devices, and other resources necessary for stand-alone operation within a frame similar to that of a tower client computer.

rack-mounted server: computer server that is designed for deployment in a standard 19 inch ICT site rack as defined by EN 60297 [i.1]

NOTE 1: Equivalent national standards include EIA-310 and DIN 41494.

NOTE 2: For the purposes of this specification, a blade server is considered under a separate product category and excluded from the rack-mounted product category.

resilient server: computer server designed with extensive reliability, availability, serviceability (RAS) and scalability features integrated in the micro architecture of the system, central processing unit and chipset

NOTE: The characteristics of a resilient server are described in the annex A.

server appliance: computer server that is bundled with a pre-installed operating system and application software which is used to perform a dedicated function or set of tightly coupled functions.

NOTE 1: Server appliances deliver services through one or more networks (e.g. IP or Storage Area Network), and are typically managed through a web or command line interface.

NOTE 2: Server appliance hardware and software configurations are customized by the vendor to perform a specific task (e.g. name services, firewall services, authentication services, encryption services, and voice-over-IP services), and are not intended to execute user-supplied software.

controller system: computer or computer server that manages a benchmark evaluation process

data averaging interval: time period over which all samples captured by the high-speed sampling electronics of the power analyzer are averaged to provide the measurement set

deployed power: total average power level at a defined load and load level of the number of servers provisioned to meet a fixed aggregate peak load.

efficiency: a defined workload output divided by the resource input to the system.

geomean: geometric mean of n items

NOTE:

$$Geomean(x) = \left(\prod_{i=1}^{n} x_i \right)^{1/n} = \sqrt[n]{x_1 \times ...... \times x_n}$$

hard disk drive: primary computer storage device which reads and writes to one or more rotating magnetic disk platters

High Performance Computing (HPC) system: computing system which is designed (or assembled), optimized, marketed and sold to execute highly parallel applications for higher performance computing applications

NOTE 1: HPC systems feature large number of clustered homogeneous nodes, clustered primarily to increase computational capability, and often featuring high speed inter-processing interconnects as well as large memory capability and bandwidth.

NOTE 2: HPC systems may be comprised of multiple clusters of homogenous nodes, for which the clusters may be heterogeneous.

hypervisor: supervisory system level software that establishes and manages a virtualized environment which enables multiple operating systems to run on a single physical system at the same time

ICT equipment: equipment providing data storage, processing and transport services
NOTE to entry: a combination of Information Technology Equipment and Network Telecommunications Equipment

ICT site: site containing structures or group of structures dedicated to the accommodation, interconnection and operation of ICT equipment together with all the facilities and infrastructures for power distribution and environmental control together with the necessary levels of resilience and security required to provide the desired service availability

idle state: operational state of a computer server in which the operating system and other software have completed loading, the computer server is capable of completing workload transactions, but no active workload transactions are requested or pending by the system (i.e., the computer server is operational, but not performing any useful work)

NOTE: For systems where ACPI standards are applicable, idle state correlates only to ACPI System Level S0.

I/O device: device which provides data input and output capability between a computer server and other devices

NOTE 1: An I/O device may be integral to the computer server motherboard or may be connected to the motherboard via expansion slots.

I/O port: physical circuitry within an I/O device where an independent I/O session can be established

NOTE: A port is not the same as a connector receptacle; it is possible that a single connector receptacle can service multiple ports of the same interface

maximum power: peak sustained or root means square power consumption value while operating the worst case functions

memory: server component external to the processor in which information is stored for immediate use by the processor

motherboard: main circuit board of the server typically accommodating the processor, memory, BIOS, expansion slots and enabling the attachment of additional circuit boards

network client (testing): computer or computer server that generates workload traffic for transmission to a unit under test connected via a network switch

network equipment: device whose primary function is to pass data among various network interfaces, providing data connectivity among connected devices (e.g. routers and switches) via the routing of data packets encapsulated according to IP, Fibre Channel, InfiniBand or similar protocol

normalised performance: relative performance values calibrated to a baseline common to the set of equipment being evaluated

power supply unit: self-contained device, physically separable from the motherboard of the computer server, that converts a.c. or d.c. input power to one or more d.c. power outputs for the purpose of powering the computer server via a removable or hard-wired electrical connection

a.c.-d.c power supply unit: power supply unit that converts line-voltage alternating current (a.c.) input power into one or more direct current (d.c.) power outputs for the purpose of powering a computer server

d.c.-d.c power supply unit: power supply unit that converts line-voltage direct current (d.c.) input power to one or more d.c. outputs for the purpose of powering a computer server

NOTE: For purposes of this specification, a d.c-d.c converter (also known as a voltage regulator) that is internal to a computer server and is used to convert a low voltage d.c (e.g. 12 VDC) into other d.c power outputs for use by computer server components is not considered a d.c-d.c power supply

single output power supply unit: power supply unit designed to deliver the majority of its rated output power to one primary direct current (d.c.) output for the purpose of powering a computer server

NOTE 1: Single-output power supply units may offer one or more standby outputs that remain active whenever connected to an input power source.

NOTE 2: The total rated power output from any additional power supply units outputs that are not primary and standby outputs shall be no greater than 20 W.

NOTE 3: Power supply units that offer multiple outputs at the same voltage as the primary output are considered single-output power supply units unless those outputs are generated from separate converters or have separate output rectification stages, or have independent current limits.

multi-output power supply unit: power supply unit designed to deliver the majority of its rated output power to more than one primary direct current (d.c.) output for the purpose of powering a computer server

NOTE 1: Multi-output power supply units may offer one or more standby outputs that remain active whenever connected to an input power source.

NOTE 2: The total rated power output from any additional power supply units outputs that are not primary and standby outputs is greater than or equal to 20 W.

processor: the central processing unit of the computer server comprising logic circuitry that responds to and processes the basic instructions that drive the server

product category: second-order classification or sub-type within a product group or form-factor that is based on product features and installed components and used in the present document to determine qualification and test requirements

reported maximum power: highest maximum power recorded on the twelve SERT™ worklet scores for the two tested configurations

server processor utilization: ratio of processor computing activity to full-load processor computing activity at a specified voltage and frequency, measured instantaneously or with a short term average of use over a set of active and/or idle cycles

server product family: high-level description for a group of servers sharing one chassis and motherboard combination that may contain multiple hardware and software configurations

solid state drive: storage device that uses memory chips instead of rotating magnetic platters for data storage

storage product: fully-functional storage system that supplies data storage services to clients and devices attached directly or through a network

NOTE 1: A storage product may be composed of integrated storage controllers, storage devices, embedded network elements, software, and other devices. While storage products may contain one or more embedded processors, these processors do not execute user-supplied software applications but may execute data-specific applications (e.g. data replication, backup utilities, data compression, install agents).

NOTE 2: Components and subsystems that are an integral part of the storage product architecture (e.g. to provide internal communications between controllers and disks) are considered to be part of the storage product.

NOTE 3: Components that are normally associated with a storage environment at the ICT site level (e.g. devices required for operation of an external Storage Area Network) are not considered to be part of the storage product.

storage server: enterprise storage device which contains the same components as a computer server together with ≥ 10 storage devices and software (vendor or 3rd party) that supports storage system connectivity, capacity optimization management, virtualized storage environment and software defined storage

NOTE: Supporting features are described by the product’s datasheet description and is either accompanied with vendor specific utilities and/or commercially available software supporting these functions.

Uninterruptible Power Supply (UPS): combination of convertors, switches, and energy storage devices (such as batteries) constituting a power system for maintaining continuity of load power in case of input power failure

weighted geometric mean (geomean): geometric mean calculated using a predetermined factor for each the elements prior to aggregation

worklet: a synthetic software routine, using real application functions focused on a particular type of computing activity, which stresses a particular characteristic of the system

NOTE: A floating point and integer performance stress code would be a CPU worklet

3.2 Symbols

For the purposes of the present document, the following symbols apply:

Not applicable.

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

a.c., AC	alternating current
ACPI	Advanced Configuration and Power Interface
BIOS	Basic Input/Output System
CENELEC	European Committee for Electrotechnical Standardization
CLC	see CENELEC
CPU	Central Processor Unit
d.c., DC	direct current
DDR	Double Data Rate
DIMM	Dual In-line Memory Module
GPGPU	General-Purpose computing on Graphics Processing Units
HDD	Hard Disk Drive
HPC	Higher Performance Computing
I/O	Input/Output
ISO/IEC	International Standards Organisation/International Electrotechnical Committee
LISN	Line Impedance Stabilization Network
PCI	Peripheral Component Interconnect
PFC	Power Factor Correction
PDU	Power Distribution Unit
PSU	Power Supply Unit
RAID	Random Array of Independent Drives
SAS	Serial Attached SCSI
SERT™	Server Efficiency Rating Tool
SPEC®	Standard Performance Evaluation Corporation
SSD	Solid State Drive
RAS	Reliability, Availability and Serviceability
RASM	Reliability, Availability, Serviceability and Manageability
rms	Root Mean Square

UPS Uninterruptible Power Supply
UUT Unit Under Test

4 Server product categories and representative product family configurations

4.1 General

To compare or evaluate systems, the evaluation or metric used must be against like products. Like products can be grouped into categories.

It is appropriate to comparisons of servers in the same category. Products of different categories shall not be compared permitted.

Even though servers are classified in categories by the type of system, each server is customized by its configuration to best match the application for which they are being sold or purchased.

As a result, a product is represented by a fixed set of configurations which bound the family of configurations sold.

For an appropriate evaluation, the category must be defined and the product family configurations itemized. The categories of servers are segmented by form-factor, number of sockets and other design characteristics that will define different groups of product with distinctively different energy profiles.

The difference in energy profile and resulting usage model is why different category products shall not be compared to each other.

The product family configuration establishes a single group representing the efficiency of the product, as this then covers the range of configurations of that product that would be sold. Since the products sold are custom-configured, there should be a minimum of 2 configurations that would represent the family of configurations, maximum performance and minimum performance configurations.

Typical configurations are optional and would be represented by the maximum and minimum performance configurations.

NOTE: For some categories of product, such as 1-socket blades, the number of configurations is limited. For categories which contain limited configurations, “typical” may end up becoming a slight variation of minimum or maximum performance configurations.

4.2 Applications and metric applicability

Computer servers are architected in such a way as to be configurable to different groups of applications. The full configuration, including logical and physical elements, is optimized to deliver the most effective means of operating those applications.

The software component of the metric is designed to execute typical real world applications and must stress the elements included in the server systems.

By stressing the elements in the fashion real world applications would, a representation of the output and its associated power demand is established.

Changes in application target, technology, configuration, or elements will impact the results and applicability of the metric. Therefore, it is necessary to establish categories of products which describe the elements which groups similar products, determine representative configurations, and ensure applicability of the metric.

For example, if the category of system relies on highly parallel graphical calculation accelerators, and the metric (typically its software) does not fully exercise this element, neither performance nor energy reported by the metric is applicable. Metrics are made applicable by augmenting the performance or energy to account for missing areas so long as the comparisons are across systems in the same category are consistent with the general grading of the system.

Categories or comparison groups of servers are formed by a combination of physical characteristics and limitations. The categories are separated base on computer architecture and physical differences that determine a different energy profile unique to that group. The server metric in this specification is targeted for general purpose servers and may not be applicable to certain categories of systems due to the elements described in clause 4.1.

4.3 Computer servers

4.3.1 General requirements

For the purposes of the present document a computer server shall meet all of the following criteria:

- be marketed and sold as a computer server;
- be designed for and listed as supporting one or more computer server operating systems (OS) and/or hypervisors;
- be designed such that all processors have access to shared system memory and are visible to a single OS or hypervisor;
- be targeted to run user-installed applications typically, but not exclusively, enterprise in nature;
- be packaged and sold with one or more a.c.-d.c or d.c-d.c. power supplies.

and shall provide support for error-correcting code and/or buffered memory (including both buffered dual in-line memory modules (DIMMs)) and buffered on board configurations).

4.3.2 Managed computer servers

For the purposes of the present document a managed computer server shall meet both of the following criteria:

- be designed to be configured with redundant power supplies;
- contain an installed dedicated management controller (e.g. service processor).

4.3.3 Form-factors

For the purposes of the present document, two computer server form-factors are considered as defined in clause 3.1:

- rack-mounted;
- pedestal.

4.4 Computer server product categories

The general purpose server metrics of the present document are applicable to the following server product categories as defined and uniquely characterised in clause 3.1:

- blade 1-socket server;
- blade 2-socket server;
- blade 2-socket resilient server;
- blade 4-socket resilient server;
- direct current server;
- pedestal 1-Socket server (unmanaged);
- rack 1-socket server;
- rack 2-socket server;
- rack 2-socket resilient server;
- rack 4-socket resilient server.

A general purpose server metrics of the present document are not applicable to the following server product categories as defined and uniquely characterised in clause 3.1:

- fully fault tolerant server;
- High Performance Computing (HPC) system;
- hyper-converged server;
- large scale server;
- multi-node server;
- network equipment;
- network server;
- server appliance;
- storage equipment;
- storage server.

4.5 Computer server configuration

4.5.1 General

A server product family configuration shall:

- be from the same model line or machine type;
- either share the same form-factor (i.e., rack-mounted, blade, pedestal) or share the same mechanical and electrical designs with only superficial mechanical differences to enable a design to support multiple form-factors;
- either share processors from a single defined processor series or share processors that plug into a common socket type;
- share the power supply unit(s).

4.5.2 “High-end” performance configuration

“High-end” performance configuration of a server product family is one of the two configurations used to represent a group of servers sold in various configurations by a manufacturer.

High-end performance configuration is comprised of two solid state drives (SSDs), processor with the highest product of core count and frequency and memory capacity (in GB) equal to 1.0 to 2.0 times the product of the number of central processing units (CPUs), cores and hardware threads that represents the highest performance product model within the product family.

4.5.3 “Low-end” performance configuration

“Low-end” performance configuration of a server product family is one of the two configurations used to represent a group of servers sold in various configurations by a manufacturer.

Low-end performance configuration is comprised of two 10,000 rpm hard disk drives (HDDs), processor with the lowest product of core count and frequency and memory capacity (in GB) equal to 0.5 to 0.75 times the product of the number of CPUs, cores and hardware threads that represents the lowest performance product model within the product family.

5 Metrics

5.1 Active state metric

5.1.1 Worklets

The SERT™ tool reports performance and power data for:

- 7 CPU worklets and of the 7 CPU worklets, 6 are used in the creation of the single value efficiency metric i.e. Compress, LU, CryptoAES, SOR, Sort, and SHA 256;

  NOTE: XML_validate is represented in the capacity metric and is therefore excluded from the CPU workload calculation.

- 2 memory worklets: Flood3 and Capacity3;
- 2 storage worklets: Sequential and Random;
- 1 hybrid worklet: Hybrid SSJ.

  NOTE: The Hybrid SSJ worklet is also considered as a CPU workload for the purposes of creating a single combined efficiency metric.

For each worklet, data is reported for a set of proportional performance intervals and associated, measured power values along with other test measurements.

An interval efficiency value can be calculated by dividing each normalized performance value by the interval power value.

The set of individually measured Performance and Power values with their associated efficiency value will be referred to as interval data.

5.1.2 Formulae

5.1.2.1 General

In order to create a single energy efficiency metric for a server it is necessary to combine the worklet interval performance and power values for all the different worklets.

The method to create a single efficiency metric uses the following general procedure:

a) combining individual worklet normalized performance and power interval data to obtain an overall Efficiency, Performance and Power value for the worklet;

b) combining worklet Efficiency, Performance and Power values by workload type (CPU, Memory, Storage) to obtain a workload type value;

c) combining worklet types using a weighted geomean to obtain a single, total server Efficiency, Performance or Power value.

The geomean function offers the best option to combine the interval data to produce a worklet efficiency score and the workload (CPU, memory, storage) worklet efficiency scores to create a workload efficiency score. Using the geomean prevents any single performance, power, worklet or workload efficiency score from unduly influencing the single metric.

5.1.2.2 Generic approach

$$Eff_{server} = {Perf_{server}}/{Pwr_{server}} \qquad \text{Equation 1}$$

$$Perf_{server} = \exp\left[W_{CPU} \times \ln\left(Perf_{CPU}\right) + W_{Memory} \times \ln\left(Perf_{Memory}\right) + W_{Storage} \times \ln\left(Perf_{Storage}\right)\right] \qquad \text{Equation 2}$$

$$Pwr_{server} = \exp\left[W_{CPU} \times \ln\left(Pwr_{CPU}\right) + W_{Memory} \times \ln\left(Pwr_{Memory}\right) + W_{Storage} \times \ln\left(Pwr_{Storage}\right)\right] \qquad \text{Equation 3}$$

Within equations 2 and 3

$$Perf_{CPU} = \left(\prod_{i=1}^{7} Perf_i\right)^{1/7} \text{ and } Pwr_{CPU} = \left(\prod_{i=1}^{7} Pwr_i\right)^{1/7} \qquad \text{Equation 4}$$

where i = 1 for worklet$_{Compress}$,
i = 2 for worklet$_{LU}$,
i = 3 for worklet$_{SOR}$,
i = 4 for worklet$_{Crypto}$,
i = 5 for worklet$_{Sort}$,
i = 6 for worklet$_{SHA256}$ and
i = 7 for worklet$_{Hybrid\ SSJ}$.

$$Perf_{Memory} = \left(\prod_{i=1}^{2} Perf_i\right)^{1/2} \text{ and } Pwr_{Memory} = \left(\prod_{i=1}^{2} Pwr_i\right)^{1/2} \qquad \text{Equation 5}$$

where i = 1 for worklet$_{Flood3}$,
i = 2 for worklet$_{Capacity3}$.

$$Perf_{Storage} = \left(\prod_{i=1}^{2} Perf_i\right)^{1/2} \text{ and } Pwr_{Storage} = \left(\prod_{i=1}^{2} Pwr_i\right)^{1/2} \qquad \text{Equation 6}$$

where i = 1 for worklet$_{Sequential}$,
i = 2 for worklet$_{Random}$.

5.1.2.3 Standard approach

The generic approach does not specifically integrate the efficiency at each of the intervals. By construct, the geomean of the interval performance is a fixed percentage of $Perf_{max}$. As a result, aggregating efficiency ratios is a mathematically equivalent and simplified aggregation method which also incorporates the interval efficiencies. The impact and verification of the application of the metric use the Deployed Power Assessment method of annex B. The simplified integrated method for power managed servers is as follows:

$$Eff_{server} = \exp\left[W_{CPU} \times \ln\left(Eff_{CPU}\right) + W_{Memory} \times \ln\left(Eff_{Memory}\right) + W_{Storage} \times \ln\left(Eff_{Storage}\right)\right] \qquad \text{Equation 7}$$

Within equation 7

$$Eff_{CPU} = \left(\prod_{i=1}^{7} Eff_i\right)^{1/7} \qquad \text{Equation 8}$$

where i = 1 for worklet$_{Compress}$,
i = 2 for worklet$_{LU}$,
i = 3 for worklet$_{SOR}$,
i = 4 for worklet$_{Crypto}$,
i = 5 for worklet$_{Sort}$,
i = 6 for worklet$_{SHA256}$ and
i = 7 for worklet$_{Hybrid\ SSJ}$.

$$Eff_{Memory} = \left(\prod_{i=1}^{2} Eff_i \right)^{1/2} \qquad \text{Equation 9}$$

where i = 1 for $worklet_{Flood3}$,
i = 2 for $worklet_{Capacity3}$.

,

$$Eff_{Storage} = \left(\prod_{i=1}^{2} Eff_i \right)^{1/2} \qquad \text{Equation 10}$$

where i = 1 for $worklet_{Sequential}$,
i = 2 for $worklet_{Random}$.

Where the worklet interval aggregation is.

$$Eff_{worklet} = \left(\prod_{i=1}^{n} Eff_i \right)^{1/n} \qquad \text{Equation 11}$$

$$Perf_{worklet} = \left(\prod_{i=1}^{n} Perf_i \right)^{1/n} \qquad \text{Equation 12}$$

$$Pwr_{worklet} = \left(\prod_{i=1}^{n} Pwr_i \right)^{1/n} \qquad \text{Equation 13}$$

where
n = the number of worklet interval values for the worklet being evaluated
Eff_i = Efficiency value of the worklet interval i
$Perf_i$ = Performance value of worklet interval i
Pwr_i = Power value of the worklet interval i

5.1.3 Weightings

Independent of the approach taken the following weighting are applied:

W_{CPU} is the weighting assigned to the CPU worklets = 0,60

W_{Memory} is the weighting assigned to the Memory worklets = 0,35

$W_{Storage}$ is the weighting assigned to the Storage worklets = 0,05

5.2 Idle state metric

Idle power is the alternating current (a.c.) power of the device with no activity after running the SERT™ worklets. Idle power is reported by SERT™. Idle power of the system typically scales based on the features and performance added. When establishing minimum criteria for server idle power, additional power allocation shall be provided consistent with the performance and added features in the device.

As inactive power levels may vary significantly as increased circuitry, features and complexity is added to the system, idle state power values become increasingly poor indicators of energy efficiency. If idle power is used as a criteria, idle power allocations shall be applied to ensure more efficient servers as measured by Deployed Power Assessments (see annex B).

6 Test setup

6.1 General

A single test setup shall be used to undertake the measurements to be used determine both the active state and idle state metrics..

A separate test setup shall be used to evaluate the power supply efficiency.

6.2 Input power

Input power shall be as specified in Table 1 and Table 2. The frequency for input power shall be as specified in Table 3.

Table 1: Input power requirements for products with nameplate rated ≤ 1500 W

Product type	Supply voltage	Voltage tolerance (%)	Total harmonic distortion (%, max.)
Servers with a.c.-d.c. single-output PSUs	230 VAC or 115 VAC (see NOTE)		
Servers with a.c.-d.c. multi-output PSUs	230 VAC or 115 VAC (see NOTE)		
Optional Testing Conditions for ac-d.c (Japanese market)	100 VAC		
Three-phase servers (North American market)	208 VAC	1,0	2,0
Three-phase servers (European market)	400 VAC		
D.C powered	48 VDC		
High Voltage DC (see ETSI EN 300 132-3-1 [2])	260 - 400 VDC		
NOTE 230 V ac refers to the European market and 115 V ac refers to the North American market			

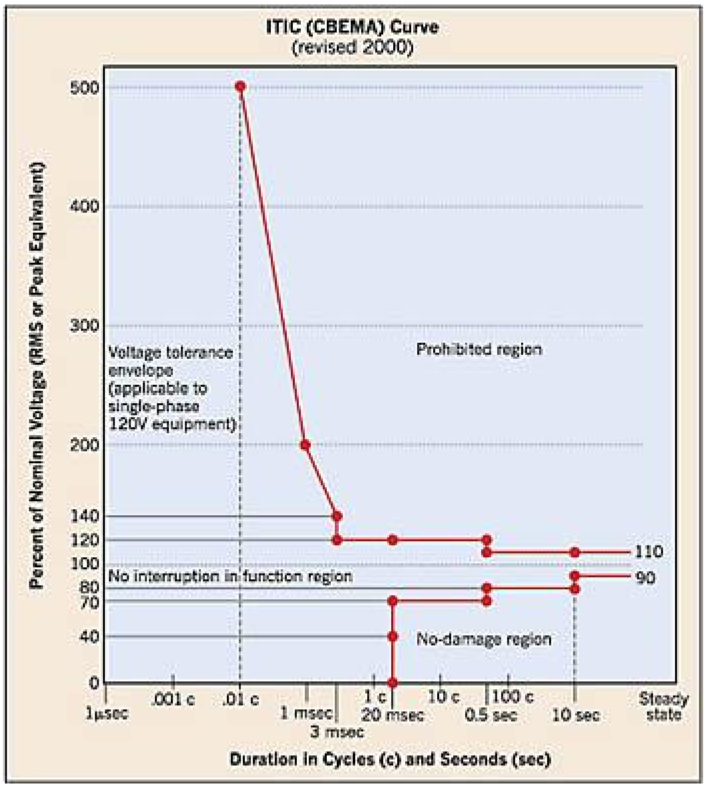
Table 2: Input power requirements for products with nameplate rated > 1500 W

Product Type	Supply Voltage	Voltage tolerance (%)	Total harmonic distortion (%, max.)
Servers with a.c.-d.c. single-output PSUs	230 VAC or 115 VAC (see NOTE)		
Servers with a.c.-d.c. multi-output PSUs	230 VAC or 115 VAC (see NOTE)		
Optional Testing Conditions for ac-d.c (Japanese mrket)	100 VAC		
Three-phase servers (North American market)	208 VAC	4,0	5,0
Three-phase servers (European market)	400 VAC		
D.C powered	48 VDC		
High Voltage DC (see ETSI EN 300 132-3-1 [2])	260 - 400 VDC		
NOTE 230 V ac refers to the European market and 115 V ac refers to the North American market			

Table 3: Input frequency requirements for all products

Supply voltage	Frequency	Frequency tolerance
100 VAC	50 Hz or 60 Hz	
115 VAC	60 Hz	
230 VAC	50 Hz or 60 Hz	1,0
Three-phase (North American Market)	60 Hz	
Three-phase (European Market)	50 Hz	

The power supply characteristics shall be not lie within the prohibited region of Figure 1.



(as published in IEEE Standard 1100:1999 [i.3]

Figure 1- ITIC (CBEMA) curve

6.3 Environmental conditions

6.3.1 Ambient Temperature:

Ambient temperature shall be 25 ± 5 °C.

6.3.2 Relative Humidity:

Relative humidity shall be between 15% and 80%.

6.4 Power analyser

The power analyzer shall report true root mean square (r.m.s.) power and at least two of the following parameters: voltage, current, and power factor.

The power analyzer shall:

a) be chosen from the list of power measuring devices specified in the most current SERT™ Design Document [3];

b) have been calibrated within a year of the test date, by a standard traceable to the National Institute of Science and Technology (USA) or a counterpart national metrology institute in other countries.

c) feature an available current crest factor of 3 or more at its rated range value;

d) for power analyzers that do not specify the current crest factor, the power analyzer shall be capable of measuring an amperage spike of at least 3 times the maximum amperage measured during any 1 second sample;

e) have a minimum frequency response of 3,0 kHz;

f) have a minimum resolution of:

- 0,01 W for measurement values less than 10 W;
- 0,1 W for measurement values from 10 W to 100 W; and
- 1,0 W for measurement values greater than 100 W.

g) have a power measurement accuracy of no greater than 1,0 %;

h) have a logging performance of:

- minimum reading rate: one set of measurements (power measurement in W) per second;
- data averaging interval equal to the reading rate

6.5 Temperature sensor

The temperature sensor shall:

a) be chosen from the list of temperature measuring devices specified in the most current SERT™ Design Document [3];

b) have a temperature measurement accuracy of no greater than ± 0.5 °C when measured no more than 50 mm in front of (upwind of) the main airflow inlet of the Unit Under Test (UUT).

c) have a logging performance of minimum reading rate: four samples per minute.

6.6 Active state test tool

The active state test tool shall be that specified in the most current version of SERT™, provided by the Standards Performance Evaluation Corporation® (SPEC) [3].

6.7 Controller system

The controller system shall be capable of the following functions:

a) start and stop each segment (phase) of the performance benchmark;

b) control the workload demands of the performance benchmark;

c) start and stop data collection from the power analyzer so that power and performance data from each phase can be correlated;

d) store log files containing benchmark power and performance information;

e) convert raw data into a suitable format for benchmark reporting, submission and validation;

f) collect and store environmental data, if automated for the benchmark.

The controller system may be a server, a desktop computer, or a laptop and shall be used to record power from the equipment specified in 6.3 and temperature data from the equipment specified in 6.4.

The controller system and the UUT shall be connected to each other via an Ethernet network switch.

6.8 General SERT requirements

Any additional requirements specified in any SPEC®, or the most current SERT™ supporting documents shall be followed, unless otherwise specified in this test method.

Supporting documents from SPEC® include:

a) SPEC Power and Performance Methodology;

b) SPEC Power Measurement Setup Guide;

c) SPEC PTDaemon Design Document;

d) SERT Design Document [3];

e) SERT Run and Reporting Rules;

f) SERT User Guide;

g) SERT JVM Options;

h) SERT Result File Fields.

7 Unit Under Test

7.1 Configuration

The configuration of the UUT shall be as specified in Table 4.

Table 4 - Configuration of UUT

A)	As-shipped Condition	Products shall be tested in their “as-shipped” configuration, which includes both hardware configuration and system settings, unless otherwise specified in this test method. Where relevant, all software options shall be set to their default condition.
B)	Measurement location	All power measurements shall be taken at a point between the a.c. power source and the UUT Uninterruptible Power Supply (UPS) units shall not be connected between the power meter and the UUT. The power meter shall remain in place until all Idle and Active State power data are fully recorded. When testing a blade system, power shall be measured at the input of the blade chassis (i.e. at the power supplies that convert ICT site distribution power to chassis distribution power).
C)	Air flow	Purposefully directing air in the vicinity of the measured equipment in a way that would be inconsistent with normal ICT site practices is prohibited
D)	Power supplies	All PSUs shall be connected and operational. For UUT with multiple PSUs: - all power supplies shall be connected to the a.c. power source and operational during the test; - if necessary, a Power Distribution Unit (PDU) may be used to connect multiple power supplies to a single source (if a PDU is used, any overhead electrical use from the PDU shall be included in the power measurement of the UUT). For blade servers with half-populated chassis configurations, the power supplies for the unpopulated power domains can be disconnected (Table 5, D) for more information)
E)	Power Management and Operating System	The as-shipped operating system or a representative operating system shall be installed. Products that are shipped without operating systems shall be tested with any compatible operating system installed. For all tests, the power management techniques and/or power saving features shall be left as-shipped. Any power management features which require the presence of an operating system (i.e. those that are not explicitly controlled by the Basic Input Output System (BIOS) or management controller) shall be tested using only those power management features enabled by the operating system by default.
F)	Storage	Products shall be tested for qualification with at least one HDD or one SSD installed. Products that do not include pre-installed drives (HDD or SSD) shall be tested using a storage configuration used in an identical model for sale that does include preinstalled drives. Products that do not support installation of drives (HDD or SSD) and, instead, rely exclusively on external storage solutions (e.g. storage area network) shall be tested using external storage solutions.

G)	Blade System and Dual/Multi-Node Servers	A blade system or dual/multi-node server shall have identical configurations for each node or blade server including all hardware components and software/power management settings. These systems shall also be measured in a way that ensures all power from all tested nodes/blade servers is captured by the power meter during the entire test.
H)	Blade Chassis	The blade chassis, at a minimum, shall have power, cooling, and networking capabilities for all the blade servers. The blade chassis shall be populated as specified in Table 5, D). All power measurements for blade systems shall be made at the input of the blade chassis.
I)	BIOS and UUT System Settings	All BIOS settings shall remain as-shipped unless otherwise specified in the test method
J)	Input/Output (I/O) and Network Connection	The UUT shall have at least one port connected to an Ethernet network switch. The switch shall be capable of supporting the UUT's highest and lowest rated network speeds. The network connection shall be live during all tests, and, although the link shall be ready and able to transmit packets, no specific traffic is required over the connection during testing. For the purpose of testing ensure the UUT offers at least one Ethernet port (using a single add-in card only if no onboard Ethernet support is offered).
K)	Energy Efficient Ethernet	Products shipped with support for Energy Efficient Ethernet (compliant with IEEE 802.3az) shall be connected only to Energy Efficient Ethernet compliant network equipment during testing. Appropriate measures shall be taken to enable EEE features on both ends of the network link during all tests.

7.2 Test procedure

The UUT test configuration shall be in accordance with Table 5.

Table 5 - Test configuration of UUT

A)	The UUT shall be tested with the processor sockets populated.
B)	The UUT shall be installed in a test rack or other static location and shall not be physically moved until testing is complete.
C)	For a multi-node system, the power consumption per node of the UUT shall be measured in the fully-populated blade chassis configuration. All multi-node servers installed in the blade chassis shall be identical, sharing the same configuration.
D)	For a blade system, the blade server power consumption of the UUT shall be measured in the half-populated blade chassis configuration with an additional option of testing the UUT in the fully populated blade chassis configuration. For blade systems, the blade chassis shall be populated as follows: 1. Individual blade server configuration All blade servers installed in the blade chassis shall be identical, sharing the same configuration (homogeneous). 2. Half -chassis population (required) The number of blade servers required to populate half the number of single-wide blade server slots available in the blade chassis shall be calculated. For blade chassis having multiple power domains, the number of power domains shall be chosen that is closest to filling half of the blade chassis. In a case where there are two choices that are equally close to filling half of the blade chassis, test with the domain or combination of domains which utilize a higher number of blade servers. Example 1: A blade chassis supports up to 7 single-wide blade servers on two power domains. One power domain supports 3 blade servers and the other supports 4 blade servers. In this example, the power domain which supports 4 blade servers would be fully populated during testing, while the other power domain would remain unpopulated. Example 2: A blade chassis supports up to 16 single-wide blade servers on four power domains. Each of the four power domains supports 4 blade servers. In this example, two of the power domains would be fully populated during testing, while the other two power domains would remain unpopulated. All user manual or manufacturer recommendations shall be followed for partially populating the blade chassis, which may include disconnecting some of the power supplies and cooling fans for the unpopulated power domains. If user manual recommendations are not available or are incomplete, then the following guidance shall be followed: Completely populate the power domains. If possible, disconnect the power supplies and cooling fans for unpopulated power domains. Fill all empty bays with blanking panels or an equivalent airflow restriction for the duration of testing. 3. Full-chassis population (optional) All available blade chassis bays shall be populated.

	All power supplies and cooling fans shall be connected. Proceed with all required tests in the test procedure as specified in Table 6
E)	The UUT shall be connected to a live Ethernet (IEEE 802.3) network switch. The live connection shall be maintained for the duration of testing, except for brief lapses necessary for transitioning between link speeds.
F)	The Controller System required to provide SERT™ workload harness control, data acquisition, or other UUT testing support shall be connected to the same network switch as the UUT and satisfy all other UUT network requirements. Both the UUT and Controller System shall be configured to communicate via the network.
G)	The power meter shall be connected to an ac voltage source set to the appropriate voltage and frequency for the test, as specified in Table 4.
H)	The UUT shall be connected to the measurement power outlet on the power meter following the guidelines in Table 4, B).
I)	The data output interface of the power meter and the temperature sensor shall be connected to the appropriate inputs of the Controller System.
J)	It shall be verified that the UUT is configured in its as–shipped configuration.
K)	It shall be verified that the Controller System and UUT are connected on the same internal network via an Ethernet network switch.
L)	Using a normal ping command, It shall be verified that the Controller System and UUT can communicate with each other.
M)	The most current SERT [4] shall be installed on the UUT and the Controller System as specified in the most current SERT™ User Guide [4]

8 Measurement

8.1 Measurement for active state

The measurement shall be in accordance with Table 6.

Table 6 - Measurement of active state

A)	The UUT shall be re-booted.
B)	Between 5 and 15 minutes after the completion of initial boot or log in, the most current SERT™ User Guide [4] shall be followed to engage SERT™
C)	All steps outlined in the most current SERT User Guide [4] shall be followed to successfully run SERT™ There shall be no manual intervention or optimization of the Controller System, UUT, or its internal and external environment during the execution of SERT™.
D)	Once SERT™ is completed, the following output files shall be included with all testing results: 1. Results.xml 2. Results.html 3. Results.txt 4. All results-chart png files (e.g. results-chart0.png, results-chart1.png, etc.) 5. Results-details.html 6. Results-details.txt 7. All results-details-chart png files (e.g. results-details-chart0.png, results-details-chart1.png, etc.)

8.2 Re-verification and audits

Due to the manufacturing variance in components, number of significant power elements in the system, and run to run variations, retest of an individual server, a server from the same family, or a server family will result in values that may vary from the initial testing.

Adjusting for a 90% confidence interval, retesting values that are within 15% of the passing level shall be acceptable under re-test. Re-test or audit resulting in greater than 15% error to the designated passing level shall be re-evaluated as a product family to determine compliance.

For audits and re-verification of Power Supply Units, re-testing of PSU efficiency shall be not be lower than the declared value by more than 2% and the power factor shall not be lower than the declared value by more than 10%. Variance beyond these levels shall require re-evaluation of the Power Supply Unit.

8.3 Measurement for power supply

8.3.1 Measurement for internal power supply

The measurement shall be in accordance with Table 7.

Table 7 - Measurement of internal power supply

A)	For all types of internal power supplies, the efficiency and the power factor shall be measured at 10%, 20 %, 50 % and 100 % of the rated [nameplate] output power
B)	For single-output internal power supplies, the efficiency shall also be measured at 10 % of the rated [nameplate] output power
C)	For single-output internal power supplies with a rated output power greater than 500 W, the power factor shall also be measured at 10 % of the rated [nameplate] output power.
D)	Test setup, test conditions, and measurement instrument specifications shall comply with clause 6.3.
E)	This test procedure assumes that the internal power supply meets following criteria: Detailed input and output ratings are available on the name plate or in manufacturer's literature, specifying the maximum loads that can safely be placed on each individual d.c output voltage bus and, where necessary, groupings of those voltage busses. The power supply has connectors that allow the d.c. output voltage busses to be connected and disconnected from the powered product non-destructively. The power supply can be easily detached from the housing of the product it powers, without causing harm to other circuits and components of the product.
F)	In the event the above criteria are not met, a test board (see 8.3.2) shall be provided to enable testing.
NOTE 1: The power supply can be easily detached from the housing of the product it powers, without causing harm to other circuits and components of the product. NOTE 2: Such data could already be available from the manufacturer of the power supply; in such cases, the manufacturer could decide to use them. However, where 3rd party test results are used, it is the responsibility of the manufacturer to assess the trustworthiness of the sources. NOTE 3: The EPRI/ECOVA Generalized Test Protocol [i.2] is an acceptable basis for providing the required data.	

8.3.2 Measurement for test board power supply

8.3.2.1 General

Tests specified in this section shall be made on the power supply of the computer under test, after it has been disconnected from the powered parts and extracted from the housing. Alternatively, another unit, representative of the built-in power supply may be used.

8.3.2.2 Test loads

Active loads such as electronic loads or passive loads such as rheostats may be used as d.c test loads. They shall be able to maintain the required current loading set point for each output voltage within an accuracy of ± 0.5%.

8.3.2.3 Test leads and wiring

Appropriate wires shall be used to avoid excessive overheating and reduce voltage drop across the wires. If measurements are not taken directly at the connector pins, voltage drop against the additional wires shall be taken into account.

NOTE: If applicable, voltage drop across the input wires will be subtracted from the measured input voltage, and the voltage drop across the output wires will be added the measured input voltage for the measurement of power efficiency.

8.3.2.4 Warm up time

Whereas internal temperature of the components could impact its efficiency, the power supply under test shall be loaded up to the test load a period of at least 15 minutes or until the reading over two consecutive five-minute intervals does not change by more than ± 0.2%.

8.3.2.5 Power measurements

The true RMS wattmeter used to carry out a.c. input power measurements shall meet the requirements of clauses 5.7 and 5.8 of EN 62623:2013 [1]. Input power shall be determined using an averaging technique over a minimum of 32 input cycles utilizing the measurement instrument averaging function.

For appliances connected to more than one phase, the power measurement instrument shall be equipped to measure the total power of all phases connected.

D.c. output power measurements shall be made with a suitably calibrated voltmeter and ammeter.

8.3.2.6 Power Factor (PF) measurement

In order to avoid interaction between the a.c. power source and a possible Power Factor Correction (PFC) circuit, a Line Impedance Stabilization Network (LISN) shall be inserted in series between the power source and the a.c. input of the power supply under test.

9 Measurement report

The following shall be listed in the measurement report:

a) the active state KPI;

b) the idle state KPI.

It is accepted that the measurement accuracy of the metrics is ± 10 %. Any subsequent assessment within this range shall be considered to be consistent with the quoted value.

The following shall also include the following:

c) automated SERT™ test report and supporting data;

d) author, site, and date of the testing;

e) product category;

f) product Family and configuration type;

g) server configuration including:

    i) system test model number;

    ii) manufacturer name;

    iii) server product ID;

    iv) component manufacturer, product ID, number of units and component product type for CPU, memory, drive (HDD or SSD);

h) Power Supply Unit test information or reference to previously conducted test report;

i) maximum inlet server test temperature prior to and during the test;

j) revision numbers for each of test software elements used;

k) Java™ revision and source used.

l) controller product model ID

m) test equipment (power meter, thermal sensor/meter): manufacturer, model, ID, and calibration date

n) SERT™ tool suite revisions

o) SERT™ product configuration revision

Annex A: Resilient server requirements (normative)

A.1 RAS features

Reliability features support a server's ability to perform its intended function without interruption due to component failures (e.g. component selection, temperature and/or voltage de-rating, error detection and correction).

Availability features support a server's ability to maximize operation at normal capacity for a given duration of downtime (e.g. redundancy - both at micro- and macro-level).

Serviceability features support a server's ability to be serviced without interrupting operation of the server (e.g. hot plugging).

A.2 RAS requirements

In order to classify as a resilient server all of the features listed in Table A.1 shall be provided.

Table A.1: RAS requirements

A)	**Processor RAS and Scalability**
A.1	Processor RAS: The processor must have capabilities to detect, correct, and contain data errors, as described by all of the following: Error detection on L1 caches, directories and address translation buffers using parity protection; Single bit error correction (or better) using ECC on caches that can contain modified data. Corrected data is delivered to the recipient (i.e., error correction is not used just for background scrubbing); Error recovery and containment by means of (1) processor checkpoint retry and recovery, (2) data poison indication (tagging) and propagation, or (3) both. The mechanisms notify the OS or hypervisor to contain the error within a process or partition, thereby reducing the need for system reboots; and (1) Capable of autonomous error mitigation actions within processor hardware, such as disabling of the failing portions of a cache, (2) support for predictive failure analysis by notifying the OS, hypervisor, or service processor of the location and/or root cause of errors, or (3) both.
A.2	The processor technology used in resilient and scalable servers is designed to provide additional capability and functionality without additional chipsets, enabling them to be designed into systems with 4 or more processor sockets. The processors have additional infrastructure to support extra, built-in processor busses to support the demand of larger systems.
A.3	The server provides high bandwidth I/O interfaces for connecting to external I/O expansion devices or remote I/O without reducing the number of processor sockets that can be connected together. These may be proprietary interfaces or standard interfaces such as PCIe. The high performance I/O controller to support these slots may be embedded within the main processor socket or on the system board.
B	**Memory RAS and Scalability**
B.1	Provides memory fault detection and recovery through Extended ECC.
B.2	In x4 DIMMs, recovery from failure of two adjacent chips in the same rank
B.3	Memory migration: Failing memory can be proactively de-allocated and data migrated to available memory. This can be implemented at the granularity of DIMMs or logical memory blocks. Alternatively, memory can also be mirrored.
B.4	Uses memory buffers for connection of higher speed processor -memory links to DIMMs attached to lower speed DDR channels. Memory buffer can be a separate, standalone buffer chip which is integrated on the system board, or integrated on custom-built memory cards. The use of the buffer chip is required for extended DIMM support; they allow larger memory capacity due to support for larger capacity DIMMs, more DIMM slots per memory channel, and higher memory bandwidth per memory channel than direct-attached DIMMs. The memory modules may also be custombuilt, with the memory buffers and DRAM chips integrated on the same card;
B.5	Uses resilient links between processors and memory buffers with mechanisms to recover from transient errors on the link.
B.6	Lane sparing in the processor-memory links. One or more spare lanes are available for lane failover in the event of permanent error.
C.	**Power Supply RAS**
	All PSUs installed or shipped with the server shall be redundant and concurrently maintainable. The redundant and repairable components may also be housed within a single physical power supply, but must be repairable without requiring the system to be powered down. Support must be present to operate the system in degraded mode when power delivery capability is

	degraded due to failures in the power supplies or input power loss.
D.	**Thermal and Cooling RAS**
	All active cooling components, such as fans or water-based cooling, shall be redundant and concurrently maintainable. The processor complex must have mechanisms to allow it to be throttled under thermal emergencies. Support must be present to operate the system in degraded mode when thermal emergencies are detected in system components.

A.3 System resilience requirements

In order to classify as a resilient server at least six of the features listed in Table A.2 shall be provided.

Table A.2: System resiliency requirements

E.	**System Resiliency**
E.1	Support of redundant storage controllers or redundant path to external storage
E.2	Redundant service processors
E.3	Redundant d.c-d.c regulator stages after the power supply outputs
E.4	The server hardware supports runtime processor de-allocation
E.5	I/O adapters or drives (HDD or SSD) are hot-swappable;
E.6	Provides end to end bus error retry on processor to memory or processor to processor interconnects
E.7	Supports on-line expansion/retraction of hardware resources without the need for operating system reboot (“on-demand” features)
E.8	Processor Socket migration: With hypervisor and/or OS assistance, tasks executing on a processor socket can be migrated to another processor socket without the need for the system to be restarted
E.9	Memory patrol or background scrubbing is enabled for proactive detection and correction of errors to reduce the likelihood of uncorrectable errors; and
E.10	Internal storage resiliency: Resilient systems have some form of RAID hardware in the base configuration, either through support on the system board or a dedicated slot for a RAID controller card for support of the server’s internal drives. F. System Scalability – All of the following shall be present in the server:
E.11	Higher memory capacity: >=8 DDR3 or DDR4 DIMM Ports per socket, with resilient links between the processor socket and memory buffers; and
E.12	Greater I/O expandability: Larger base I/O infrastructure and support a higher number of I/O slots. Provide at least 32 dedicated PCIe Gen 2 lanes or equivalent I/O bandwidth, with at least one x16 slot or other dedicated interface to support external PCIe, proprietary I/O interface or other industry standard I/O interface

Annex B: Deployed Power Assessment (informative)

B.1 Overview

This clause provides an overview of the Deployed Power Assessment methodology for the validation of a proposed server efficiency metric

In order to evaluate potential metrics, a series of assessments are made of the relative power impact that selecting the server offers to a definitive confirmation of a metric's ability to predict efficiency of the server as deployed in the ICT site.

The assessment comprises a graphical correlation and rank comparison between the power impact of a number of servers deployed to execute a defined workload versus the aggregated efficiency metric for the server. The selection of servers mimics the provisioning method used by IT professionals when determining the data processing needs of that ICT site. This is termed the Deployed Power Assessment method.

The Deployed Power Assessment is based on determining the ability of provisioning a set of servers for a targeted workload that results in a minimum expenditure of energy at the ICT site, across the various utilization levels. The metric validation is to ensure that a better efficiency score will result in a lower deployed server power demand to execute that work in a ICT site. The use of a deployed power calculation enables differentiation between the effectiveness of a low performance, low power server and a high performance, high power server, as it enables an assessment of the number of servers and their associated energy use required to deliver a given workload in a ICT site or office environment.

To calculate the number of servers needed to perform a workload and their associated deployed power, it is necessary to select a workload level to use to calculate the number of deployed servers.

1) Target workload should represent the maximum composite of the work targeted for the ICT site. Based on actual workloads deployed in ICT sites, the weights used are 60% CPU, 35% Memory and 5% Storage. The performance of each sub-workload category, e.g. CPU, Memory, and Storage, is the geomean of the maximum performance of each worklet in the category

2) A value of "100 x the maximum performance value of the group of tested servers" defines the target workload performance level for the evaluation. This value is large enough avoid quantization effects.

3) For evaluation of the resulting workload, three assessment workloads were selected: a CPU intensive workload, a memory intensive workload and a weighted workload which mimics the workload weighting of the combined metric being assessed.

4) Additionally, the following power/utilization/workload types to be used in our comparison analysis. The intent is to assess a combination of workload types and power consumption levels that a set of servers will experience in an operating environment:

   a) idle power as measured by the SERT™ tool;

   b) geomean of the power for all workloads at the 25% utilization (light workload);

   c) geomean of the power for all workloads at the 50% utilization (medium workload);

   d) geomean of the power for all workloads at the 100% utilization (heavy workload);

   e) weighted deployed power using the metric weighting;

   f) weighted deployed power using the CPU intensive workload weightings (85% CPU, 15% memory);

   g) weighted deployed power using the Memory intensive workload weightings (40% CPU, 60% memory).

Multiple workloads and power use scenarios were assessed to validate that the combined metric is balanced and representative of efficiency across the range of workloads that servers are expected to perform and to avoid the assertion that use of a single power/workload/utilization might be biased to a particular outcome.

B.2 Determining the number of deployed servers

B.2.1 General

Any attempt at a single efficiency metric based upon the SERT™ tool must use some method to combine the individual worklet values to create a single value. The aggregation used to determine a single value is with the geomean combinatory method and the designated component (CPU, memory and storage) weightings to calculate the number of servers required.

In order to determine the number of servers required for any given server model one must determine both a performance target for the dataset and a performance capability for each individual server. The individual worklet performance values reported in the SERT™ tool are combined to determine the number of servers required to meet the performance target.

B.2.2 Establishing target performance

In order to minimize quantization issues, since deployed power is based on an integral number of servers, a 100 times the maximum performance of the highest performance server in the data set for the performance target. The number of servers required to meet a desired performance level is calculated according to Equation B.1.

$$Deployed_{QTY_n} = Roundup\left(\frac{100 \times Perfmax_{allservers}}{Perfmax_n}\right)$$ Equation B.1

where:

$Perfmax_{AllServers}$ = the performance values for all servers in the data set
$Perfmax_n$= the maximum performance of server n.

B.2.3 Weighting factors

A weighted geomean (CPU 60 %: Memory 35 %: Storage 5 %) of performance, designated as the weighted performance maximum, is used.

Equation B.2 aggregates the geomean of the normalized maximum performance values to obtain a single maximum performance number for each individual server. The weighted maximum performance number represents the maximum capability of the server and what would be used to provision a ICT site.

$$Perfmax_{weighted} = \exp\left[0,6 \times \ln\left(Perfmax_{CPU}\right) + 0,35 \times \ln\left(Perfmax_{Memory}\right) + 0,05 \times \ln\left(Perfmax_{Storage}\right)\right]$$ Equation B.2

$$Deployed_{QTY_n} = Roundup\left(\frac{100 * Max(Perf_{All\ Servers})}{Perf_n}\right)$$ where:

$Perfmax_{\text{weighted}}$ represents the newly calculated weighted maximum performance
$Perfmax_{\text{worklet}}$ represents the maximum performance score from the SERT™ data base.

And:

$$Perfmax_{CPU} = \left(\prod_{i=1}^{7} Perfmax_i\right)^{1/7}$$ Equation B.3

where i = 1 for $\text{worklet}_{Compress}$,
i = 2 for worklet_{LU},
i = 3 for worklet_{SOR},
i = 4 for worklet_{Crypto},
i = 5 for worklet_{Sort},
i = 6 for worklet_{SHA256} and
i = 7 for $\text{worklet}_{Hybrid\ SSJ}$.

$$Perfmax_{Memory} = \left(\prod_{i=1}^{2} Perfmax_i\right)^{1/2}$$ Equation B.4

where i = 1 for worklet_{Flood3}, i = 2 for $worklet_{Capacity3}$.

$$Perfmax_{Storage} = \left(\prod_{i=1}^{2} Perfmax_i \right)^{1/2}$$ Equation B.5

where i = 1 for $worklet_{Sequential}$,
i = 2 for $worklet_{Random}$.

Bibliography

Final Report - Ecodesign Technical Assistance Study on Standards for Lot 9 Enterprise Servers and Enterprise Data Storage

Server efficiency Metric from SERT Worklet Results DIGITALEUROPE Proposal, 5 August 2016, Version 2.0 http://www.digitaleurope.org/DesktopModules/Bring2mind/DMX/Download.aspx?Command=Core_Download&EntryId=2334&language=en-US&PortalId=0&TabId=353

ISO/IEC 30134-4: Information technology – ICT sites – Key Performance Indicators - Part 4: IT Equipment Energy Efficiency for Servers ($ITEE_{SV}$)

Deployed power, The Green Grid,

Energy Star for servers

ASHRAE TC9.4

ASHRAE TC9.9, Data Center Equipment Thermal Design Guidelines

CBEMA (Computer Business Equipment Manufacturers Association) Curves, ITI Council, 2007

European Environmental Citizens Organisation for Standards www.ecostandard.org

SPEC Power and Performance Methodology;

SPEC Power Measurement Setup Guide https://www.spec.org/power/docs/SPECpower-Measurement_Setup_Guide.pdf

SPEC PTDaemon Design Document;

SERT Run and Reporting Rules;

SERT JVM Options;

SERT Result File Fields.

https://www.spec.org/power/docs/SPEC-Power_Analyzer_Acceptance_Process.pdf

TGG, TGG/SPEC creating a single value metric {placeholder for real title}

History

Document history		
<Version>	<Date>	<Milestone>
V0.0.1	04/10/2016	First draft
V0.0.2	08/02/2017	After face to face meeting
V0.0.3	08/11/2016	Working draft V.0.0.2 with correct format etc).
V0.0.4	30/04/2017	Working draft following comment process
V0.0.5	31/05/2017	Pre-stable draft
V0.0.6	19/06/2017	Stable draft as submitted to TC EE and ***editHelp!***