ETSI TR 104 004 V1.1.1 (2024-09)

Environmental Engineering (EE);
Processor power management functionality of servers

Reference
DTR/EE-EEPS60

Keywords
server

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 - APE 7112B
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° w061004871

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The declarations pertaining to these essential IPRs, if any, are 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 (https://ipr.etsi.org/).

Pursuant to the ETSI Directives including the ETSI IPR Policy, no investigation regarding the essentiality of IPRs, 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.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

DECT™, **PLUGTESTS™**, **UMTS™** and the ETSI logo are trademarks of ETSI registered for the benefit of its Members. **3GPP™** and **LTE™** are trademarks of ETSI registered for the benefit of its Members and of the 3GPP Organizational Partners. **oneM2M™** logo is a trademark of ETSI registered for the benefit of its Members and of the oneM2M Partners. **GSM**® and the GSM logo are trademarks registered and owned by the GSM Association.

Foreword

This Technical Report (TR) has been produced by ETSI Technical Committee Environmental Engineering (EE).

Modal verbs terminology

In the present document "**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.

Executive summary

CPU Power Management functions (P-states and C-states or their equivalent) offer hardware available, software initiated functions to reduce CPU voltage and frequency when workload demands are low or absent to reduce server energy use. Depending on the amount of time a server operates below 25 % utilization or is idle, with longer periods generating higher energy savings, power demand can be reduced by up to 50 % with the full implementation of P-states and C-states. These power demand reductions improve SERT overall efficiency scores by up to 30 %. Power management functions come at a cost of reduced server performance and increased response time (exit latency). Data from SERT measurements on a single configuration with power management turned on resulted in performance reductions at the 100 % utilization level in the Hybrid ssj worklet of 6 % for an EPYC™ CPU and 14 % for a Power9™ CPU compared to power management off. For response time, the literature shows that P-state transition times and C-state exit latencies cause response delays that are problematic for some workloads and applications. As an example, Operating System (OS) managed power management profiles can interfere with virtualization programs like VMware® impeding performance rates on the virtualized images. Power management can be a benefit or a problem depending on the data centre operations, workloads and applications. While the implementation of a power management profile can be beneficial in many instances, the optimum P-state and C-state settings and the choice of controlling software - BIOS, hypervisor or operating system - will depend on the specific use case. In some cases, such as high-speed financial trading, network providers or high-performance computing, power management functions will need to be turned off to ensure the performance and response times demanded by those workloads.

Introduction

Data centre energy consumption and IT equipment efficiency are increasingly the focal point of data centre customers and operators, non-governmental organizations and government regulators with an interest in data centre operations. There are a host of options and approaches to reduce data centre energy consumption that can focus on the individual IT equipment, the data centre IT system, or the facility equipment. As servers are the majority of the IT equipment in a data centre and have a validated energy efficiency metric, the SPEC SERT suite, server energy efficiency requirements have attracted the most attention of regulatory efforts to improve data centre energy efficiency.

Server processor power management features, which can moderate power demand based on factors such as CPU utilization, provide a means for server manufacturers and/or data centre operators to reduce server power consumption and improve server energy efficiency. In many CPU architectures, the primary means of server processor power management is through implementation of CPU P-states and C-states. Enablement of power management states can reduce performance and increase response time. The present document will focus on the characteristics, benefits and limitations of CPU power management features in deployment situations.

Server power management features are also available for system fans, memory, storage, Graphic Processing Unit and I/O components. Their combined implementation introduces greater complexity and requires a higher level of integration with the server's platform firmware and operating system. Component power management is beyond the scope of the present document. More specifically the present document will provide a brief overview of CPU P-states and C-states, SERT test data detailing the power demand reductions achieved by different power management implementations, the SERT measured active state energy efficiency improvements, and the performance and exit latency impacts of server power management features. The present document only addresses the current state of these technologies.

1 Scope

The present document is focused on addressing the characterization of the process power management functionality of servers.

The processor power management of servers is limited to those within scope of Commission Regulation (EU) 2019/424 [i.1].

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative 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.

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

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] Commission Regulation (EU) 2019/424 of 15 March 2019 laying down ecodesign requirements for servers and data storage products pursuant to Directive 2009/125/EC of the European Parliament and of the Council and amending Commission Regulation (EU) No 617/2013.

[i.2] C. Chou, L. N. Bhuyan and D. Wong: "µDPM: Dynamic Power Management for the Microsecond Era", 2019 IEEE International Symposium on High Performance Computer Architecture (HPCA), Washington, DC, USA, 2019, pp. 120-132, doi: 10.1109/HPCA.2019.00032.

[i.3] I. Curtis: "Hot Chips 2020 Live Blog, Next Gen Intel Xeon Ice Lake-SP", Hot Chips Conference, August 17, 2020.

[i.4] D. Molka, R. Schone, M. Werner: "Wake-up latencies for processor idle states on current x86 processors", Computer Science - Research and Development, Vol. 30, doi: 10.1007/s00450-014-0270-z; 2014/04/01.

[i.5] S. Kanev, K. Hazelwood, G. Wei and D. Brooks: "Tradeoffs between power management and tail latency in warehousescale applications", 2014 IEEE International Symposium on Workload Characterization (IISWC), Raleigh, NC, 2014, pp. 31-40, doi: 10.1109/IISWC.2014.6983037.

[i.6] Broadcom®: "Virtual machine application runs slower than expected in ESXi".

[i.7] Lenovo™: "Tuning VMware for Increased Performance - Lenovo ThinkSystem and x86 Servers".

[i.8] Huawei: "Huawei V5 Server Best Practice with VMware ESXi System 03".

[i.9] HPE: "Workload profiles and performance options | UEFI System Utilities User Guide for HPE Compute servers".

[i.10] The Green Grid: "SERT™ Active Efficiency: Demonstrating How SERT™ Active Efficiency Testing Includes Server Idle".