HPE ProLiant Server Power Management

Red Hat Enterprise Linux 6.x and 7.x

Contents

Technical white paper

Abstract

Power management is crucial to data center power provisioning. This document provides a brief overview of the processor-based power-saving features supported on HPE ProLiant servers, and the power management features such as Power Regulator, Power Capping, and Collaborative Power Control that are embedded in ProLiant systems. This document also discusses how these features are used and their relationship to the Red Hat® Enterprise Linux® 6.x and 7.x operating systems, including new features available with ProLiant Gen8 and later Intel®-based servers running Red Hat Enterprise Linux 6.x and 7.x.

Introduction

The Red Hat Enterprise Linux 6.x and 7.x operating systems running on HPE ProLiant servers use processor-based features to achieve better power efficiency. These processor-based features include:

- Performance states (P-states) define a set of fixed operating frequencies and voltages, where P0 represents the highest operating frequency and voltage. You can save power by entering P-states with lower frequency and voltage levels. Either the platform firmware or the operating system controls the P-states.
- Power states (C-states), excluding the C0 state, represent idle states and determine the power consumed when a processor is idle. C0 is a non-idle state with higher C-states representing idle conditions with increasing power savings. The operating system controls the C-states.
- Throttle states (T-states) define a set of fixed frequency percentages that can be used to regulate the power consumption and the thermal properties of the processor. ProLiant systems can reserve the use of T-states for the system firmware.

In addition, ProLiant servers are also capable of using the various processor states to support innovative power management features that are operating system independent and are implemented in the hardware and firmware:

- HPE Power Regulator provides a facility to efficiently control processor power usage and performance, either statically or dynamically, depending on the mode selected.
- HPE Power Capping allows an administrator to limit the power consumed by a server.
- HPE Dynamic Power Capping offers the additional feature of ensuring that the power limit set by an administrator is maintained by reacting to a spike in server workload more rapidly than basic HPE Power Capping.

The Power Regulator and Power Capping technologies are designed to work in conjunction with each other. To make the operating system aware of Power Capping, Hewlett Packard Enterprise provides the Collaborative Power Control technology. This is a two-way communication mechanism established between the operating system and platform firmware; it can be used by the operating system and hardware collaboratively to choose the appropriate performance level for the server. Support for this technology is present in Red Hat Enterprise Linux 6.x and 7.x, along with ProLiant Gen8 and newer servers.

HPE Power Regulator

HPE Power Regulator is a configurable processor power-usage feature that allows you to choose from several options for: (1) enabling the server to manage P-states, or (2) delegating control of regulating P-states to the operating system.

HPE Power Regulator is implemented within the firmware on both Intel-based and AMD-based ProLiant servers.[1] ProLiant servers provide the following HPE Power Regulator modes, which you can select from the ROM Based Setup Utility (RBSU) or through HPE Integrated Lights-Out 4 (iLO 4):

HPE Dynamic	The firmware is capable of managing the P-states. However, when the Collaborative Power Control (CPC) setting is enabled in RBSU, the OS and the firmware collaborate to attain the desired frequency for a processor. When CPC is disabled, this mode allows the firmware to exclusively control the P-states of a processor to match the server load. On HPE ProLiant Gen8 and newer servers, HPE Dynamic is the default mode with the CPC setting enabled.
HPE Static Low	The firmware controls the P-states. The P-state of the processor is static, and it is set to the P-state that corresponds to the lowest operating frequency supported by the processor.
HPE Static High	The firmware controls the P-states. The P-state of the processor is static, and it is set to P0, which corresponds to the highest operating frequency supported by the processor.
OS Control	The Red Hat Enterprise Linux 6.x and 7.x operating systems control P-states and manage the P-states according to the policy set by the administrator via the OS.

For the HPE Static Low and HPE Static High modes described above, HPE recommends that you disable CPC to ensure that the firmware has exclusive control of the P-states. CPC is located within the Advanced Power Management Options in RBSU. Disabling CPC causes Red Hat Enterprise Linux 6.x and 7.x to report in the /var/log/messages file and in the dmesg output that CPU frequency scaling is not utilized on the server.

The OS Control mode allows the ProLiant server firmware to delegate management of P-states to the Red Hat Enterprise Linux 6.x and 7.x operating systems. For general purpose workloads where a balance of high performance and power efficiency is desired, OS Control mode is recommended.

You can adjust the Power Regulator Settings through the RBSU or the HPE iLO 4 interface as shown in figure 1. You must reboot the system to change the transitions to and from the OS Control mode, but you can change the system between the other three modes dynamically.

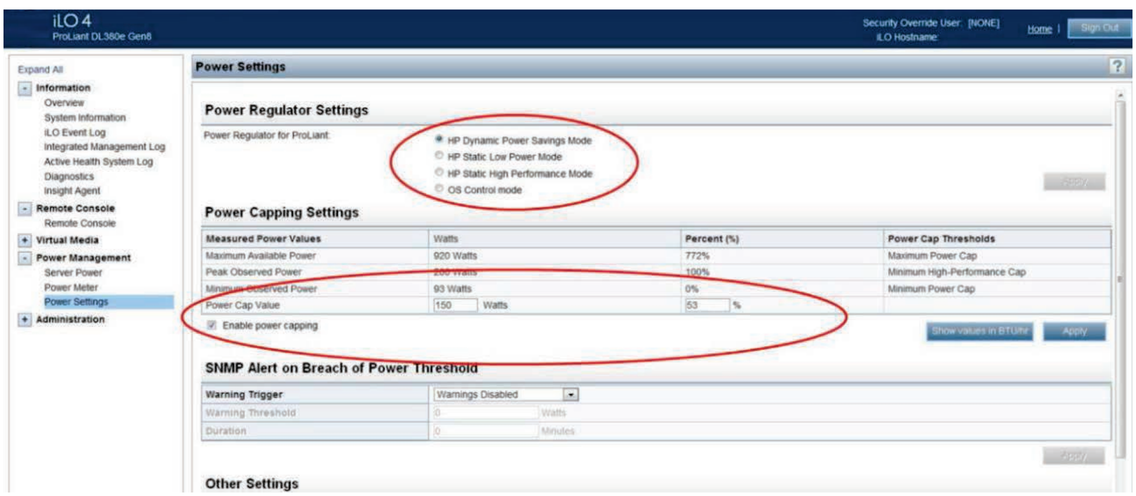

Figure 1. Configuring Power Regulator and Power Capping settings via HPE iLO 4

[1] For detailed information on HPE Power Regulator support across the different generations of ProLiant platforms, refer to: h20566.www2.hpe.com/hpsc/doc/public/display?sp4ts.oid=5219994&docLocale=en_US&docId=emr_na-c03334051

To adjust the CPC setting, you must access RBSU as shown in figure 2. Modifying this setting requires a system reboot for the setting to take effect.

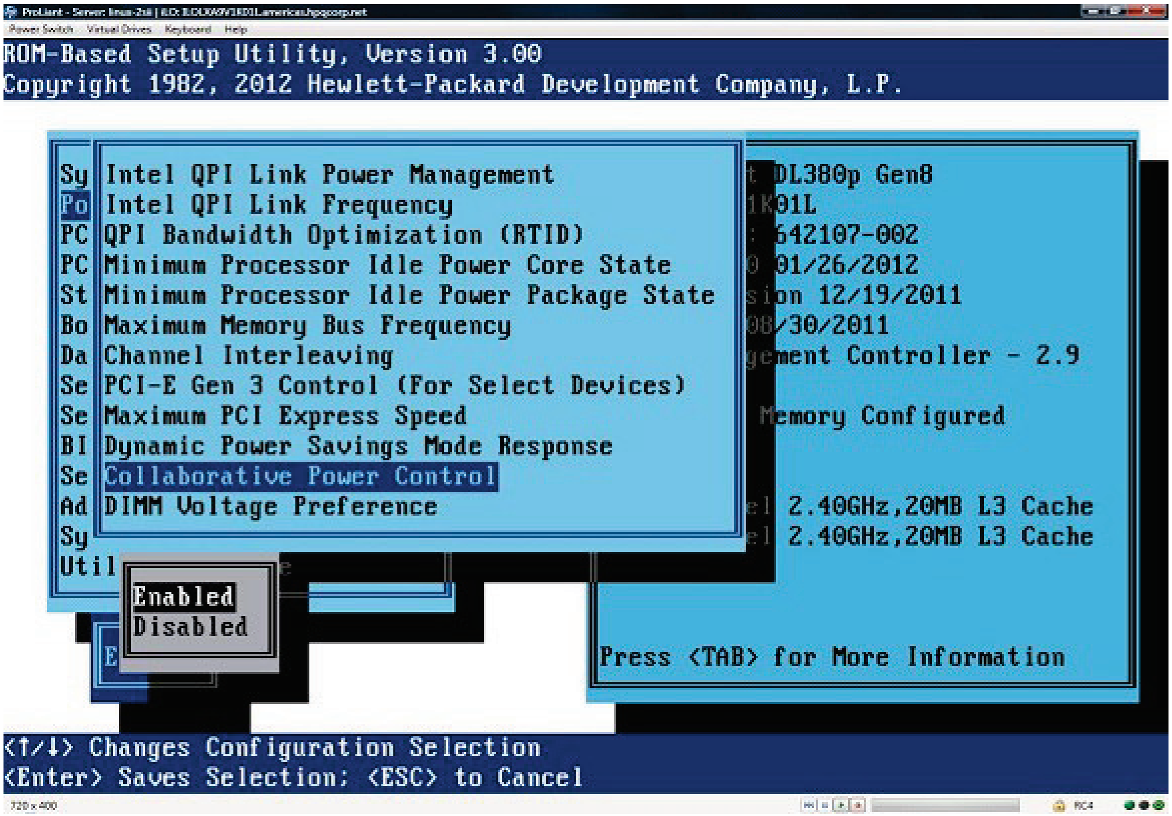

Figure 2. Configuring the Collaborative Power Control setting via RBSU

For more information on HPE Power Regulator technology, refer to: h20566.www2.hpe.com/hpsc/doc/public/display?docId=emr_na-c00300430

HPE Power Capping

HPE Power Capping satisfies data center power provisioning requirements by allowing the data center administrator to provide a power budget to a single server or a group of servers. The ProLiant platform enforces that limit by changing the processor P-states and T-states in an operating system-independent manner. HPE Power Capping is independent of the HPE Power Regulator setting. When server power is being capped under OS Control mode, the firmware overrides the power management instructions from the operating system for the duration of the capping.

As shown in figure 1, you can use HPE iLO 4 to configure a power cap. HPE iLO displays important information about maximum available power for the power supply, the peak observed power, and the minimum observed power for the server. With this information, you can select an appropriate power cap by specifying the absolute maximum watts or a percentage of the maximum observed power of the server.

For an in-depth presentation on the HPE Power Capping technology, refer to: h20566.www2.hpe.com/hpsc/doc/public/display?docId=emr_na-c01549455

Power monitoring with HPE iLO 4

HPE iLO 4 is designed to monitor current power consumption along historical timelines. As shown in figure 3, HPE iLO 4 displays the current power consumption, as well as the peak and average power consumptions for the past 24-hour and 20-minute time periods.

Figure 3. HPE iLO 4 power meter readings for 24-hour and 20-minute time periods

For more information on HPE iLO management technology, refer to: hp.com/go/ilo

Power capping demonstration with HPE iLO 4

This section of the document demonstrates the HPE Power Capping functionality by increasing the workload on a ProLiant server under the Red Hat Enterprise Linux 6.1 operating system. Figure 4 displays the HPE iLO configuration setting for a server where the capping threshold is set to 180 watts. This setting means the maximum power consumption will be limited to approximately 180 watts. Figure 5 shows that when power capping is not set, the power consumption of a server increases when the workload increases. You can also see that the maximum power consumption was 227 watts, while the minimum power consumption was 209 watts in the past five minutes. The average power consumption was 213 watts. Figure 6 demonstrates that when power capping is enabled, the server limits power consumption to 180 watts—even when the workload increases—to satisfy the power budget set by the user.

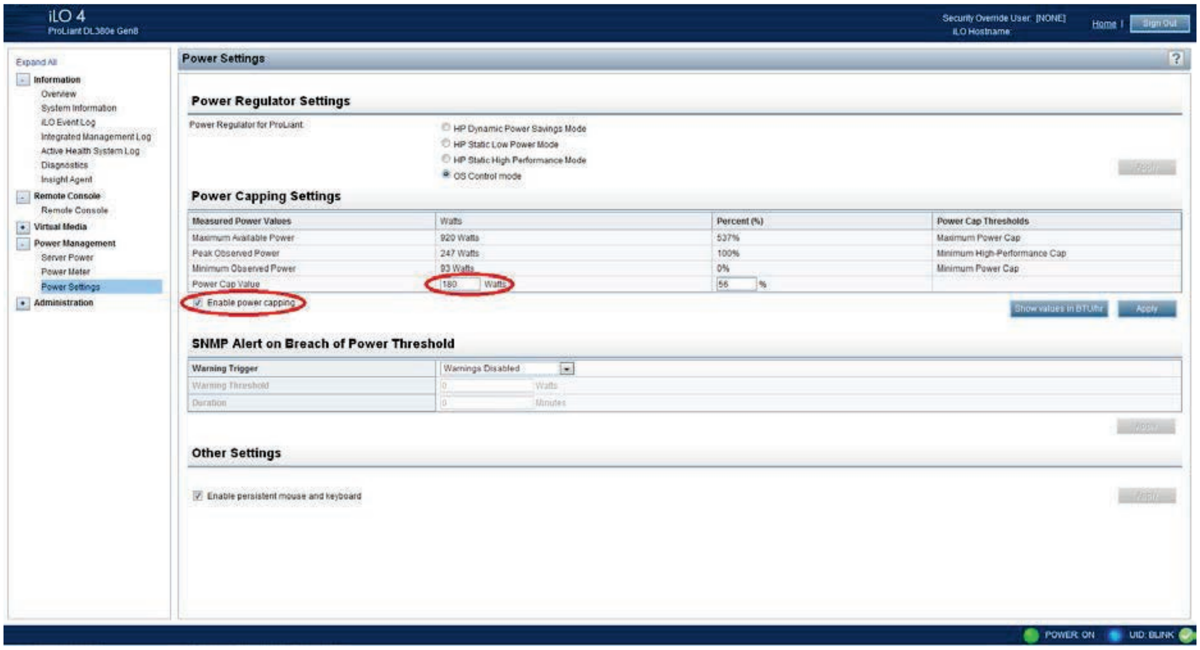

Figure 4. HPE Power Capping threshold configuration

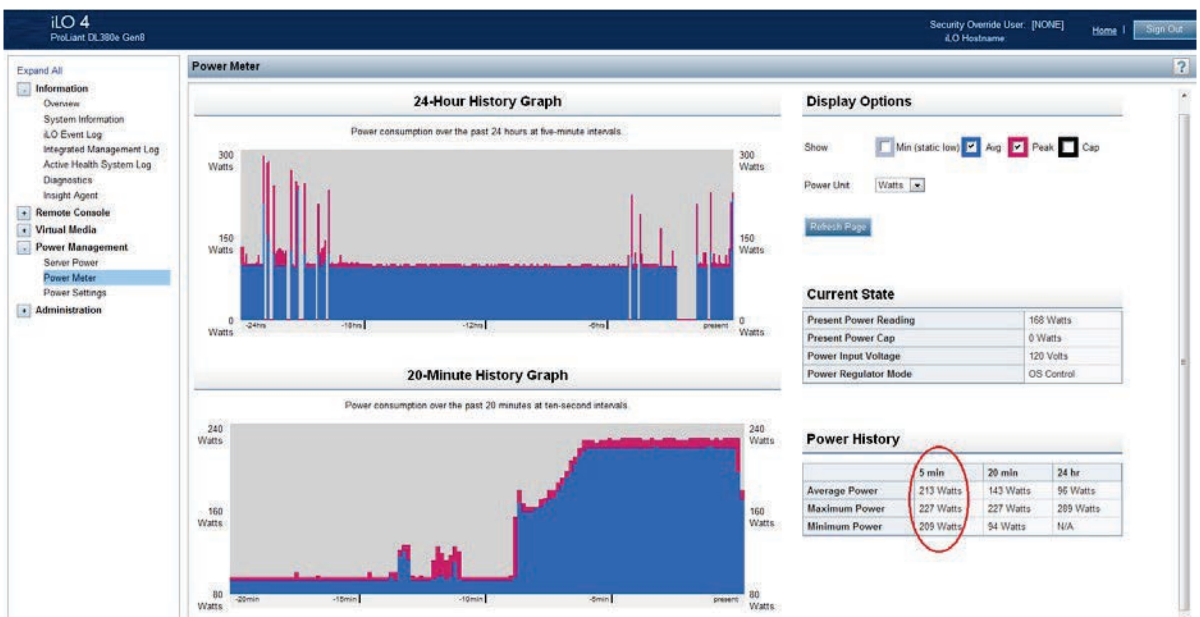

Figure 5. Power consumption without capping

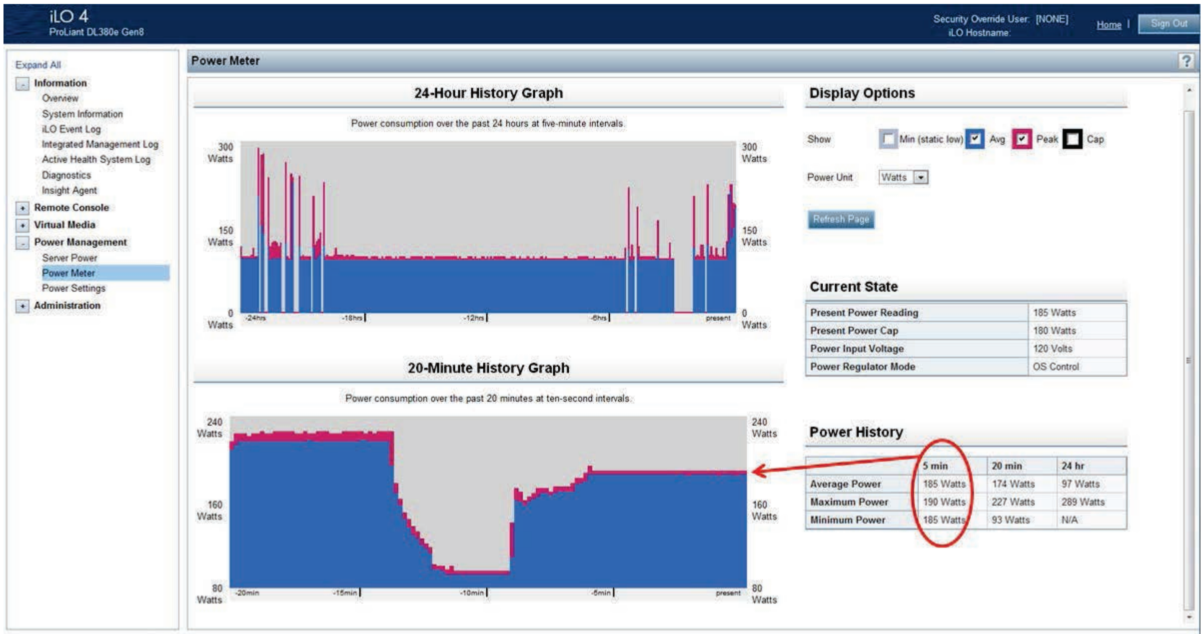

Figure 6. Power consumption with capping

ProLiant power management with Red Hat Enterprise Linux 6.x and 7.x

Red Hat Enterprise Linux 6.x and 7.x manage the power usage of ProLiant servers by adjusting the processor P-states when the HPE Power Regulator setting in RBSU is configured in OS Control mode. Typically within the Linux operating system, a governor dictates the policy, while the actual P-state transition is accomplished by a suitable P-state driver. Red Hat Enterprise Linux 6.x and 7.x offer a choice of governors, with each one implementing a different policy. The default governor for “acpi-cpufreq” P-state driver is the on-demand governor, which dynamically adjusts the processor P-states to match the load on the server. The other governors are:

- Userspace, which enables the user space program (cpuspeed)
- Directly configuring the processor frequency
- Performance, which selects the P-state corresponding to the highest supported frequency

The default governor for “intel_pstate” P-state driver is “performance”, which selects the P-state corresponding to the highest processor frequency. The other governor is “powersave”, which selects the P-state according to the system load.

On Intel-based ProLiant platforms, Red Hat Enterprise Linux 6.x and 7.x natively support the Intel Demand Based Switching with Enhanced Intel SpeedStep® technology. On AMD-based ProLiant platforms, Red Hat Enterprise Linux 6.x supports AMD’s PowerNow! technology.

The following table lists the P-state drivers on Intel-based and AMD-based ProLiant G6, G7, and newer platforms under OS Control mode.

Processor family	**P-state driver (RHEL6.x)**	**P-state driver (RHEL7.x)**
Intel® Xeon®	acpi-cpufreq	intel_pstate
AMD Opteron	powernow-k8	acpi-cpufreq

For Red Hat Enterprise Linux 6.x and 7.x to manage the processor's power consumption, the firmware must communicate information about the processor P-states and their associated frequencies to the OS. You can find this information in the file and directories under /sys/devices/system/cpu. Included in the Red Hat Enterprise Linux 6.x media is the cpufreq-info command (installed via the cpufrequtils rpm package). Red Hat Enterprise Linux 7.x does not include this utility and instead uses the cpupower utility, available in the cpupower rpm. This utility provides information about the P-states of the processors in the system in a user-friendly format. When used without arguments, cpufreq-info displays information about all processor cores (including the P-state driver), the frequency range supported by the processor, the available frequency steps (which are actually the P-states), the available and current governors, and the current frequency. Example 1 shows how cpufreq-info also supports options to display information specific to a CPU.

Example 1. Output for CPU 0 in OS Control mode (RHEL6.x)

```
# cpufreq-info -c 0
cpufrequtils 007: cpufreq-info (C) Dominik Brodowski 2004-2009
Report errors and bugs to cpufreq@vger.kernel.org, please.
analyzing CPU 0:
  driver: acpi-cpufreq
  CPUs which run at the same hardware frequency: 0
  CPUs which need to have their frequency coordinated by software: 0
  maximum transition latency: 10.0 us.
  hardware limits: 1.20 GHz - 2.00 GHz
  available frequency steps: 2.00 GHz, 2.00 GHz, 1.90 GHz, 1.80 GHz,
  1.70 GHz, 1.60 GHz, 1.50 GHz, 1.40 GHz, 1.30 GHz, 1.20 GHz
  available cpufreq governors: ondemand, userspace, performance
  current policy: frequency should be within 1.20 GHz and 2.00 GHz.
                  The governor “ondemand” may decide which speed to use
                  within this range.
  current CPU frequency is 1.20 GHz.
```

Example 2. Output for CPU 0 in OS Control mode (RHEL7.x)

```
# cpupower -c 0 frequency-info
analyzing CPU 0:
   driver: intel_pstate
   CPUs which run at the same hardware frequency: 0
   CPUs which need to have their frequency coordinated by software: 0
   maximum transition latency: 0.97 ms.
   hardware limits: 1.20 GHz - 2.50 GHz
   available cpufreq governors: performance, powersave
   current policy: frequency should be within 1.20 GHz and 2.50 GHz.
                   The governor "performance" may decide which speed to use
                   within this range.
current CPU frequency is 1.95 GHz (asserted by call to hardware).
boost state support:
   Supported: yes
   Active: yes
```

You can dynamically change the governor used under the OS Control mode by modifying the value in the `/sys/devices/system/cpu/cpu*/cpufreq/scaling_governor` file for each CPU. Red Hat Enterprise Linux 6.x provides the `cpufreq-set` command to select the governor. For Red Hat Enterprise Linux 7.x, use the cpupower command. For more information about the `cpufreq-info, cpufreq-set` and `cpupower` commands, refer to the man pages.

Collaborative power control with Red Hat Enterprise Linux 6.x and 7.x

When ProLiant servers are under OS Control mode for power management, power capping can still be imposed by the platform without knowledge of the operating system. First introduced on Intel-based ProLiant G6 servers, and included in all Gen8 and newer ProLiant servers, OS Control mode enables the server and the OS to collaborate on power management. Hewlett Packard Enterprise provides the Collaborative Power Control mechanism, which is capable of providing capping-related feedback to the operating system; it can also collaborate with the operating system to manage the power consumption of a server. This combination provides the quick response time of HPE Dynamic Power Savings, as well as delivers correct processor power information to the operating system.

CPC uses the Processor Clocking Control (PCC) interface, which coordinates processor performance between the platform firmware and the operating system. The PCC interface, jointly developed by Hewlett Packard Enterprise and Microsoft®, is publicly available—which means other platform vendors can implement it. For more information on PCC, refer to: acpica.org/sites/acpica/files/Processor-Clocking-Control-v1p0.pdf.

Platform firmware releases for Intel-based ProLiant G6 servers from August 2009 onward include support for Collaborative Power Control.

When a CPC-enabled server is configured in HPE Dynamic mode, the firmware does not present P-state information to the operating system. Instead, the firmware presents the minimum and maximum frequencies the processor supports, enabling the OS to choose any frequency within that range, rather than restricting the processor to specific P-states. If the processor is capped at that time for any reason, then the platform firmware informs the OS that the request was denied due to capping. When capping is not configured, the PCC driver continues to function in lieu of the P-state driver.

Example 3 shows a sample output for CPU 0 for a ProLiant Gen8 server running under Red Hat Enterprise Linux 6.1. In this example, notice that the driver is `pcc-cpufreq`. Only the minimum and maximum frequency limits display. Unlike under OS Control, there are no preset frequency steps.

Example 3. Output for CPU 0 in HPE Dynamic mode with CPC enabled (RHEL6.x)

```
# cpufreq-info -c 0
cpufrequtils 007: cpufreq-info (C) Dominik Brodowski 2004-2009
Report errors and bugs to cpufreq@vger.kernel.org, please.
analyzing CPU 0:
  driver: pcc-cpufreq
  CPUs which run at the same hardware frequency: 0
  CPUs which need to have their frequency coordinated by software: 0
  maximum transition latency: 0.00 ms.
  hardware limits: 1.20 GHz - 2.00 GHz
  available cpufreq governors: ondemand, userspace, performance
  current policy: frequency should be within 1.20 GHz and 2.00 GHz.
                  The governor "ondemand" may decide which speed to use
                  within this range.
  current CPU frequency is 1.20 GHz.
```

The lack of preset frequency steps can affect the output of other tools as well. For example, the PowerTOP utility displays P-state (frequency) information only when the HPE Power Regulator is configured to OS Control mode.

Idle power states (C-states) with Red Hat Enterprise Linux 6.x and 7.x

Processor power use at idle is a crucial factor in determining a server's power consumption when there is no workload to execute. Typically, when a processor has no work to perform, the operating system places the processor in a halt state signified as C1. Newer-generation processors support deep C-states (C2–C6, where C6 is the deepest state), allowing Red Hat Enterprise Linux 6.x and 7.x to take advantage of these states. The deeper the C-state, the more power the processor can save. Although C-states can significantly reduce power consumption, the drawback of going to a deeper C-state is the latency associated with the time it takes for the processor to wake up and resume executing instructions. Information about the C-states for system processors is available in `/sys/devices/system/cpu/cpu*/cpuidle/state*`.

Note
You can configure the server to not utilize the idle C-states by choosing the `No C-states` setting in RBSU.

Additional Red Hat Enterprise Linux 6.x and 7.x power management features

Red Hat Enterprise Linux 6.x and 7.x provide a comprehensive set of features for managing the power usage of ProLiant servers.

The "Green IT" features introduced in Red Hat Enterprise Linux 6.0[2] offer a range of kernel and user-space features to manage server power consumption. With the "tickless when idle" kernel feature, it is possible to reduce the number of wakeups per second from 1,024 to typically less than 30. For instance, in figure 7, notice that the "Wakeups-from-idle per second" is below 23.

Additional tools are available for monitoring system power consumption. For example, using the PowerTOP[3] tool (`powertop-1.11-4.el6.i686.rpm`), you can identify the processes responsible for waking up a processor from its idle state, and thereby drive up power consumption. You can refer to the PowerTOP documentation for further information on what the output of PowerTOP represents, and learn tips and tricks on how to best tune your server for maximum power savings.

[2] Red Hat Enterprise Linux 6: Green Computing Features
[3] An introduction to PowerTOP: 01.org/powertop

Figure 7 displays the PowerTOP v1.1 screen output on an idle 1P ProLiant DL360e Gen8 Server with Intel Xeon CPU E5-2420 Processor and 2 GB system memory running under Red Hat Enterprise Linux 6.1. The average residency in the deepest supported C-state[4] is about 49 ms. This value is due to the processor being awakened 23 per second times from its idle state. The output listing is for a case where the Intelligent Platform Management Interface (IPMI) service has been stopped on the server.[5]

The latest PowerTOP v2.0[6] (released in May 2012) provides many new features. For example, enhanced diagnostic capabilities are available by using the perf subsystem of the Linux kernel. You can monitor any of five different views by selecting one of the tabs at the top of the screen: **Overview, Idle stats, Frequency stats, Device stats**, or **Tunables**. Figure 8 displays the PowerTOP v2.0 screen output for the same system hardware used in the example of figure 7 running under Red Hat Enterprise Linux 6.3.

```
Cn                  Avg residency       P-states (frequencies)
C0 (cpu running)          ( 0.2%)       Turbo Mode     1.0%
polling             0.0ms ( 0.0%)         1.91 Ghz     0.0%
C1 mwait            0.4ms ( 0.1%)         1.80 Ghz     0.0%
C3 mwait           48.8ms (99.7%)         1400 Mhz     0.0%
                                          1200 Mhz    99.0%

Wakeups-from-idle per second : 22.8     interval: 15.0s
no ACPI power usage estimate available

Top causes for wakeups:
  65.3% ( 94.6)     <kernel core> : hrtimer_start_range_ns (tick_sched_timer)
   9.1% ( 13.1)     <kernel core> : hrtimer_start (tick_sched_timer)
   3.5% (  5.0)           cmasm2d : hrtimer_start_range_ns (hrtimer_wakeup)
   2.8% (  4.0)     <kernel core> : usb_hcd_poll_rh_status (rh_timer_func)
   1.7% (  2.5)       <interrupt> : eth3-TxRx-0
   1.7% (  2.4)       <interrupt> : uhci_hcd:usb3, hpilo
   1.4% (  2.0)     <kernel core> : clocksource_watchdog (clocksource_watchdog)
   1.1% (  1.6)       <interrupt> : ahci
   1.1% (  1.6)          events/6 : worker_thread (igb_watchdog)
   0.8% (  1.1)            hpsmhd : hrtimer_start_range_ns (hrtimer_wakeup)
   0.7% (  1.0)           hp-asrd : hrtimer_start_range_ns (hrtimer_wakeup)
   0.4% (  0.6)     <kernel core> : __enqueue_rt_entity (sched_rt_period_timer)
   0.4% (  0.5)       <interrupt> : eth0-TxRx-0
   0.4% (  0.5)       <interrupt> : eth1-TxRx-0
   0.4% (  0.5)       <interrupt> : eth2-TxRx-0
   0.3% (  0.5)         events/11 : queue_delayed_work (delayed_work_timer_fn)
   0.3% (  0.5)         events/10 : queue_delayed_work (delayed_work_timer_fn)
   0.3% (  0.5)          events/9 : queue_delayed_work (delayed_work_timer_fn)
   0.3% (  0.5)          events/8 : queue_delayed_work (delayed_work_timer_fn)
   0.3% (  0.5)          events/7 : queue_delayed_work (delayed_work_timer_fn)
   0.3% (  0.5)          events/6 : queue_delayed_work (delayed_work_timer_fn)
   0.3% (  0.5)          events/5 : queue_delayed_work (delayed_work_timer_fn)
   0.3% (  0.5)          events/4 : queue_delayed_work (delayed_work_timer_fn)
   0.3% (  0.5)          events/3 : queue_delayed_work (delayed_work_timer_fn)
   0.3% (  0.5)          events/2 : queue_delayed_work (delayed_work_timer_fn)
   0.3% (  0.5)          events/1 : queue_delayed_work (delayed_work_timer_fn)
   0.3% (  0.5)          events/0 : worker_thread (igb_watchdog)
   0.3% (  0.5)          events/0 : queue_delayed_work (delayed_work_timer_fn)
   0.3% (  0.5)          cmahostd : __nf_conntrack_confirm (death_by_timeout)
```

Figure 7. PowerTOP v1.11 output on an idle ProLiant DL360e Gen8 Server running under Red Hat Enterprise Linux 6.1 with no IPMI service

Note

You can see the P-states (frequencies) information in the PowerTOP output only when the HPE Power Regulator is configured to OS Control mode.

[4] ACPI C3 actually corresponds to hardware C6 state, which is the deepest C-state supported by the processors on that platform.

[5] Halting the IPMI driver results in a user losing the ability to remotely monitor the server. If the IPMI service is stopped to save power, it is possible to resume the IPMI service with the service ipmi restart command.

[6] PowerTOP v2.0: 01.org/powertop/blogs/ceferron/2012/powertop-v2.0-release

```
PowerTOP v2.0     Overview   Idle stats   Frequency stats   Device stats   Tunables


            Package |            Core    |            CPU 0       CPU 6
                    |                    | Actual      1217 MHz    1227 MHz
1.91 GHz      0.0%  | 1.91 GHz     0.0%  | 1.91 GHz     0.0%        0.0%
1.80 GHz      0.0%  | 1.80 GHz     0.0%  | 1.80 GHz     0.0%        0.0%
1.71 GHz      0.0%  | 1.71 GHz     0.0%  | 1.71 GHz     0.0%        0.0%
1.60 GHz      0.0%  | 1.60 GHz     0.0%  | 1.60 GHz     0.0%        0.0%
1500 MHz      0.0%  | 1500 MHz     0.0%  | 1500 MHz     0.0%        0.0%
1400 MHz      0.0%  | 1400 MHz     0.0%  | 1400 MHz     0.0%        0.0%
1300 MHz      0.0%  | 1300 MHz     0.0%  | 1300 MHz     0.0%        0.0%
1200 MHz      0.2%  | 1200 MHz     0.1%  | 1200 MHz     0.1%        0.0%
Idle         99.8%  | Idle        99.9%  | Idle        99.9%      100.0%

                    |            Core    |            CPU 1       CPU 7
                    |                    | Actual      1226 MHz    1230 MHz
                    | 1.91 GHz     0.0%  | 1.91 GHz     0.0%        0.0%
                    | 1.80 GHz     0.0%  | 1.80 GHz     0.0%        0.0%
                    | 1.71 GHz     0.0%  | 1.71 GHz     0.0%        0.0%
                    | 1.60 GHz     0.0%  | 1.60 GHz     0.0%        0.0%
                    | 1500 MHz     0.0%  | 1500 MHz     0.0%        0.0%
                    | 1400 MHz     0.0%  | 1400 MHz     0.0%        0.0%
                    | 1300 MHz     0.0%  | 1300 MHz     0.0%        0.0%
                    | 1200 MHz     0.0%  | 1200 MHz     0.0%        0.0%
                    | Idle       100.0%  | Idle       100.0%      100.0%

                    |            Core    |            CPU 2       CPU 8
                    |                    | Actual      1225 MHz    1232 MHz
                    | 1.91 GHz     0.0%  | 1.91 GHz     0.0%        0.0%
                    | 1.80 GHz     0.0%  | 1.80 GHz     0.0%        0.0%
                    | 1.71 GHz     0.0%  | 1.71 GHz     0.0%        0.0%
                    | 1.60 GHz     0.0%  | 1.60 GHz     0.0%        0.0%
                    | 1500 MHz     0.0%  | 1500 MHz     0.0%        0.0%
                    | 1400 MHz     0.0%  | 1400 MHz     0.0%        0.0%
                    | 1300 MHz     0.0%  | 1300 MHz     0.0%        0.0%
                    | 1200 MHz     0.1%  | 1200 MHz     0.0%        0.1%
                    | Idle        99.9%  | Idle       100.0%       99.9%

                    |            Core    |            CPU 3       CPU 9
                    |                    | Actual      1229 MHz    1230 MHz
                    | 1.91 GHz     0.0%  | 1.91 GHz     0.0%        0.0%
                    | 1.80 GHz     0.0%  | 1.80 GHz     0.0%        0.0%
                    | 1.71 GHz     0.0%  | 1.71 GHz     0.0%        0.0%
                    | 1.60 GHz     0.0%  | 1.60 GHz     0.0%        0.0%
                    | 1500 MHz     0.0%  | 1500 MHz     0.0%        0.0%
```

Figure 8. PowerTOP v2.0 output on an idle ProLiant DL360e Gen8 Server running under Red Hat Enterprise Linux 6.3 with no IPMI service

To support the performance-monitoring functionality, the subsystem uses the general-purpose performance counters instead of the fixed-function performance counters. The `perf` utility in Red Hat Enterprise Linux 6.x and 7.x monitors performance by selecting the monitored events. For details about performance monitoring events, refer to the Intel specification.[7]

Example 4 shows output from the `perf` utility. The command in the example specifies three events to be monitored: UnHalted Core Cycles (r003c), UnHalted Reference Cycles (r013c), and Instruction retired (r00c0). The example proves that the `perf` subsystem works correctly even if BIOS occupied the fixed-function performance counters. In this example, the `stress`[8] utility is used to make the processor as busy as possible.

[7] Intel 64 and IA-32 Architectures Software Developer's Manual Volume 3B: System Programming Guide, Part 2: download.intel.com/products/processor/manual/253669.pdf
[8] Linux-based workload generator—stress, people.seas.harvard.edu/~apw/stress/

Example 4. Output from the `perf` utility

```
# perf stat -e r003c -e r013c -e r00c0 stress --cpu 50 --timeout 120s
stress: info: [5693] dispatching hogs: 50 cpu, 0 io, 0 vm, 0 hdd
stress: info: [5693] successful run completed in 120s
  Performance counter stats for 'stress --cpu 50 --timeout 120s':
  2,715,777,736,549 r003c
    142,982,564,816 r013c
  3,031,663,572,695 r00c0
      120.004665335 seconds time elapsed
```

If you are concerned about the fixed-function performance counters being occupied by BIOS, you can disable the "Processor Power and Utilization Monitoring" option in RBSU by following these steps (the GUI is shown in figure 9):

1. Press CTRL-A.
2. Select **Service Options**.
3. Disable the **Processor Power and Utilization Monitoring** option.
4. Reboot the system.
5. After rebooting, the firmware bug message no longer appears.

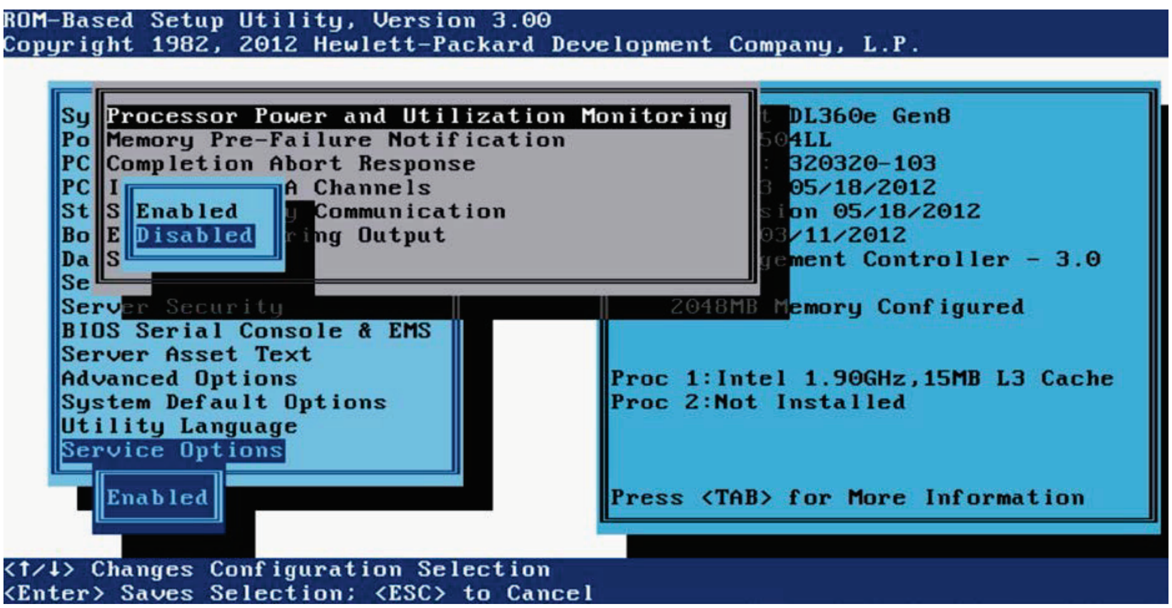

Figure 9. Disabling the Processor Power and Utilization Monitoring option in RBSU

Summary

HPE ProLiant servers are designed to save power when under load and when idle. The processor-based power management features supported in the hardware are enabled by the firmware automatically. They are also used in close coordination between the server's firmware and the Red Hat Enterprise Linux 6.x and 7.x operating systems. Typically, you do not have to activate these features; they are enabled by default.

In addition, HPE ProLiant servers include several innovative features for advanced power saving and budgeting, including HPE Power Regulator, HPE Power Capping, HPE Dynamic Power Capping, and Collaborative Power Control.

For more information

For additional information, please refer to the resources listed below.

Resource description	Web address
HPE ProLiant Gen8 servers	www8.hp.com/us/en/products/proliant-servers/index.html#!view=grid&page=1&facet=Gen8
HPE ProLiant Gen9 servers	www8.hp.com/us/en/products/proliant-servers/index.html#!view=grid&page=1&facet=Gen9
HPE Linux documentation	www8.hp.com/us/en/products/servers/solutions.html?compURI=1501074#tab=TAB4
HPE Power Capping and HPE Dynamic Power Capping	h10032.www1.hp.com/ctg/Manual/c01549455
Power Regulator for ProLiant servers	h20566.www2.hpe.com/hpsc/doc/public/display?docId=emr_na-c00300430
AMD PowerNow! Technology	support.amd.com/TechDocs/24404a.pdf
Enhanced Intel SpeedStep Technology and Demand Based Switching on Linux	software.intel.com/en-us/articles/enhanced-intel-speedstepr-technology-and-demand-based-switching-on-linux
Linux **cpufreq** kernel documentation	kernel.org/doc/Documentation/cpu-freq/
Linux **cpuidle** kernel documentation	kernel.org/doc/Documentation/cpuidle/
Intelligent Platform Management Interface	intel.com/design/servers/ipmi/
HPE Integrated Lights-Out Management Controller	hp.com/go/ilo
An introduction to PowerTOP	01.org/powertop
PowerTOP v2.0	01.org/powertop/blogs/ceferron/2012/powertop-v2.0-release
Red Hat Enterprise Linux 6: Green computing features	redhat.com/en/about/blog/red-hat-expands-green-computing-features-in-red-hat-enterprise-linux-6
Linux on HPE ProLiant servers	www8.hp.com/us/en/products/servers/solutions.html?compURI=1501074#tab=TAB2
Processor Clocking Control Interface Specification	acpica.org/sites/acpica/files/Processor-Clocking-Control-v1p0.pdf
Kernel documentation on the Linux PCC implementation	kernel.org/doc/Documentation/cpu-freq/pcc-cpufreq.txt
Introduction to PCC as presented at the Linux Foundation Collaboration Summit (LFCS) 2010	events.linuxfoundation.org/slides/lfcs2010_garbee.pdf
Intel 64 and IA-32 Architectures Software Developer's Manual Volume 3B: System Programming Guide, Part 2	intel.com/Assets/en_US/PDF/manual/253669.pdf
Linux-based workload generator—`stress`	people.seas.harvard.edu/~apw/stress/
RHEL6 Power Management Guide	access.redhat.com/documentation/en-US/Red_Hat_Enterprise_Linux/6/html/Power_Management_Guide
RHEL7 Power Management Guide	access.redhat.com/site/documentation/en-US/Red_Hat_Enterprise_Linux/7/html/Power_Management_Guide

Next steps

Hewlett Packard Enterprise welcomes your feedback. To make comments and suggestions about product documentation, send a message to: techdocs_feedback@hpe.com. Include the document title and part number if available in your message. All submissions become the property of Hewlett Packard Enterprise.

Sign up for updates

4AA5-4240ENW, September 2016, Rev. 1

Hewlett Packard Enterprise