Power Regulator for ProLiant servers

Technology brief, 5th edition

Introduction

Enterprise servers can place large demands on facility power and cooling resources. In some cases, facility power and cooling constraints can delay IT purchases. HP Power Regulator gives you a tool to increase server efficiency and free up cooling and power resources. This lets you use power/cooling resources where you need them the most. It lets you manage processor power consumption and system performance to meet your business needs.

Processor P-states

Server processors from Intel and AMD have performance state registers that can switch a processor between different performance states (P-states). Through HP Power Regulator technology, HP firmware or the OS can access and change the P-states.

Changing the performance state (the processor frequency and voltage) directs processors to operate at different power levels. This gives you some control over total system power consumption. Tables 1 and 2 show examples of P-states for Intel® Xeon® and AMD™ Opteron™ processors.

Table 1. P-states of the Quad-core Intel Xeon 5670 2.93-GHz processor

P-state	Description	Core Frequency
Pmax	Maximum performance	2.93 GHz
Pmin	Minimum power	1.6 GHz

Table 2. P-states of the AMD Opteron 6174 2.2-GHz processor

P-state	Description	Core Frequency
Pmax	Maximum performance	2.2 GHz
Pmin	Minimum power	800 MHz

Power Regulator technology

HP Power Regulator is an OS-independent power management feature of HP ProLiant servers. It is included on all HP ProLiant servers (200-series and above).

Processor power and performance state registers exposed by processor vendors let you use Power Regulator to control processor power usage and performance. Power Regulator directly adjusts the frequency and core voltage of ProLiant server processors.

Using Power Regulator, you can configure a server to maximize performance, maximize power savings, or match processor power consumption dynamically as system load changes. This lets you maintain an optimal balance of performance and power utilization under all operating conditions.

HP Power Regulator modes

You can configure Power Regulator for any of four operating modes:

• HP Static High Performance mode

- HP Static Low Power mode
- HP Dynamic Power Savings mode
- OS Control mode

Figure 1 shows the methods for controlling P-states in the multiple Power Regulator operating modes. HP has developed server ROM firmware that implements the Static High and Low modes, Dynamic Power Saving mode, and OS mode. Upgrades to the OS or application do not affect Power Regulator. The ROM firmware or the OS with a processor driver directly control processor P-states.
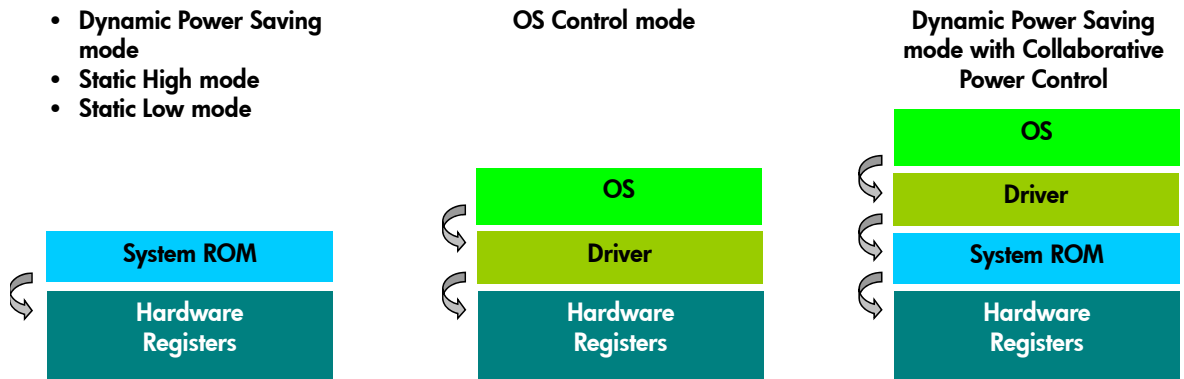
Figure 1. Methods for controlling processor P-states

Figure 1 also shows the new Collaborative Power Control (CPC) mode (described later in the paper), a capability that lets the Power Regulator control P-states by using both system ROM firmware and OS power management capabilities together. Operating systems supporting CPC can communicate with the System ROM and collaborate on performance state changes. Unlike OS Control mode, CPC ensures that the HP System ROM and OS actively manage power, regardless of whether the OS supports power management features.

Processor drivers

Processor drivers (shown in Figure 1) enable the OS Control mode and Collaborative Power Control mode. Processor drivers used in Microsoft Windows® operating systems provide an abstraction layer between specific processor designs and the OS. Windows uses information obtained from the CPU Identification instruction to determine processor type and feature set in order to load the correct processor driver.

HP Static High Performance mode

The system's processors operate continuously at the highest power and performance state (Pmax) in Static High Performance mode. Neither the system's firmware nor the OS will ever program the processors to run in a lower power or performance state. This mode is important for applications in which performance is critical and power consumption is less important.

HP Static Low Power mode

The system's processors operate continuously at the lowest power and performance state (Pmin) in Static Low Power mode. This mode is useful for environments where power availability is constrained and it is critical to lower the server's maximum power use.

HP Dynamic Power Savings mode

The Dynamic Power Savings mode is similar in concept to Intel's Demand Based Switching with SpeedStep Technology and to AMD's PowerNow!™, but it adds a significant HP innovation. Power Regulator uses an HP ROM-based algorithm to monitor processor activity. It adjusts processor performance states to match performance levels to application load. In Dynamic Power Savings mode, Power Regulator determines the amount of time each logical processor in the system spends in the operating system's idle loop. In the simplest case, where only two P-states are available, the logic is straightforward. It evaluates the following ratio where clocks per unit time is the frequency:

clocks performing useful work
total clocks available in Pmax

When the ratio is high (more time spent in the idle loop), the algorithm sets the processor to Pmin. When the algorithm detects a low ratio—indicating a high application load—Power Regulator switches the processor in real time to Pmax.

Beginning with G6 servers, Power Regulator Dynamic Power Savings mode takes advantage of all available processor P-states. In this case, the algorithm is more complicated than a simple ratio. The result is that Power Regulator will find the optimal performance state based on accurate use of Pmax and Pmin and every P-state in between.

HP ProLiant servers enable HP Dynamic Power Savings mode by default to provide significant, out-of-the-box power savings. And, it does this without significantly affecting system performance.

Measuring CPU utilization

Power-state decisions made in Dynamic Power Savings mode take into account all processor activity (kernel-mode and user-mode). Power Regulator determines CPU utilization by reading a performance event counter residing within the processor. The event counter collects non-halted clock cycles. Modern operating systems execute a HLT instruction when a processor is idle, preventing the processor from running in an idle loop while searching for work.

Processor granularity

The ROM-based dynamic power algorithm executes on every logical processor and hyperthread in the physical processor package. This means every core in a multi-core physical processor package. The power algorithm requests the higher P-state only for logical processors that require it. It tells the processor hardware to set the P-state of each logical processor based on that logical processor's current utilization.

Architectural dependencies within the physical processor may require some or all logical processors within the processor package to be set at the same P-state. When an architectural dependency is present, we allow the processor package (hardware processor management) to manage those issues. The processor package will never choose a P-state lower than the performance requirement of any logical processor in the package.

Dynamic Power Savings mode response times

Starting with ProLiant G6 servers, you are able to choose between Dynamic Power Savings mode response times (slow and fast). Power Regulator Dynamic Power Savings mode switches between P-states when the logic identifies an increase or a decrease in performance requirements. The switch occurs during a specific, user-selectable time interval. Slow and fast response times describe the fixed time interval during which the logic can identify a change in performance requirements.

Slow response times (15 seconds) are typically more power efficient, but they increase latencies and slow performance reaction to a load increase. Fast response times (0.125 seconds) react to load increases more quickly and reduce performance latencies, but often at the cost of power efficiency.

You can use this option to choose the response time that is best for the performance demands of your environment.

Benefits of Dynamic Power Savings mode

In Dynamic Power Savings mode, Power Regulator monitors and adjusts each logical processor in a system independently (per core). Dynamic Power Savings mode lets the processors operate in low power state or high power state as needed. Many customers find benefit in simply enabling the Dynamic Power Savings mode as a default setting. In HP Dynamic Power Savings mode, the Power Regulator algorithm monitors application and processor loading up to eight times per second and adjusts the P-state as needed. Optionally, you can configure the system to monitor at a slower rate using the Dynamic Power Savings mode response times (slow and fast) option in the ROM-based setup utility.

OS Control mode

In OS Control mode, the HP system ROM firmware creates the required Advanced Configuration and Power Interface (ACPI) tables. These tables let the OS support Intel's Demand Based Switching or AMD's PowerNow! power management feature. For this mode, you need to configure the OS to activate the OS-based power management feature. If you select OS Control mode and the OS does not support dynamic power management, or if you do not configure the feature through the OS, the processor will run in its highest power and performance state.

You need to reboot the system when changing into or out of OS Control mode. Changes between all other modes, controlled by HP system ROM firmware, take effect immediately.

Collaborative Power Control mode

CPC is available on Intel-based HP ProLiant G6 and later servers with supported operating systems (at the time of publication, Windows Server 2008 R2 supports CPC). Prior to CPC, firmware-based power management and OS-based power management were mutually exclusive. You could use the two resources separately, but not together. CPC takes full advantage of hardware, system firmware, and OS-based power management to choose the most efficient power usage.

CPC uses a direct, two-way interface between Microsoft Windows Server 2008 R2 and HP ProLiant Power Regulator residing in system ROM firmware. The ROM firmware receives input from the OS and iLO. CPC uses this communication pathway to gather information about choosing the appropriate performance level. Windows Server 2008 R2 has a unique set of power management features and makes requests into the ProLiant system ROM firmware to set the appropriate processor clock frequency.

CPC is an industry-standard implementation and requires:

- HP Power Regulator
- iLO 2 (or greater) management processor
- HP ProLiant system ROM firmware, which hosts the Power Regulator support.
- Supported operating systems

CPC is available only in the Power Regulator Dynamic Power Savings mode, and supported ProLiant servers enable CPC by default. You typically don't need to make adjustments to activate CPC because Dynamic Power Savings mode and CPC are default settings. Should you want to change CPC settings, the RBSU has an advanced option to "enable or disable" the CPC.

A more detailed discussion of CPC is available in "HP ProLiant G6 Server Power Management on Microsoft Windows Server 2008 R2" at http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA2-5004ENW.pdf.

Managing Power Regulator configuration

You can configure Power Regulator modes for individual server and multi-server setups using different methods.

Individual server setup

You can set Power Regulator modes using the ROM Based Setup Utility (RBSU). You typically do this on initial system boot when you have access to the RBSU menu. You must reboot to activate a Power Regulator mode configured through the RBSU.

You can also manage Power Regulator modes through the iLO browser-based interface and the iLO command-line interface (CLI). The iLO Power Management Settings page allows you to view and control the Power Regulator modes. If you prefer, iLO CLI gives you command line access to the same functions available through the iLO browser-based interface.

With the exception of the OS Control mode, Power Regulator modes configured through iLO do not require a reboot and are effective immediately. You must have the "Configure iLO Settings" privilege to change the Power Regulator mode.

Multi-server setup

iLO scripting employs iLO CLI to incorporate batch processes in a multi-server setup. You can find more information on iLO CLI and iLO scripting in the "HP Integrated Lights-Out Management Processor Scripting and Command Line Resource Guide" at http://h20000.www2.hp.com/bc/docs/support/SupportManual/c00294268/c00294268.pdf.

You can also use the Rapid Deployment Pack (RDP) for multi-server setup. RDP is another method for broadcasting configuration settings to multiple servers and requires a reboot to activate those settings.

HP Insight Control communicates with the iLO management processor on supported ProLiant servers and allows multi-server configuration of the Power Regulator modes. A multi-server setup using Insight Control and iLO does not require a reboot. Settings are effective immediately, except for OS Control mode changes, which become effective on the next reboot.

Using Power Regulator with VMware ESX

HP Power Regulator uses processor P-states to regulate server power consumption in various workload environments. In HP Dynamic Power Savings mode, Power Regulator continuously adjusts the processor power consumption to match the server workload. The server uses only the power it needs. Unfortunately, this can cause system applications to overstate overall server utilization because the measurements include data from throttled-down processors.

VMware's Distributed Resource Scheduler (DRS) and Distributed Power Manager (DPM) monitor server utilization across a pool of servers. They use these measurements to optimize workloads on host servers and can move VMs from busy host servers to idle ones. Certain applications that are very sensitive to processor latency may show less than expected performance when you use ESX and Dynamic Power Savings mode together. You may need to turn off these power management features by choosing Power Regulator Static High Performance mode in order for ESX and server hardware to achieve the best performance for such applications.

Note that selecting Static High Performance mode usually causes the system to use more power, especially when it is lightly loaded. Most applications benefit from the power savings offered by Dynamic Power Savings mode with little or no impact on performance. Therefore, if choosing Static High Performance mode does not increase performance, we recommended that you re-enable Dynamic Power Savings mode to reduce power use.

If you use the HP Power Profile pre-sets available through the RBSU in some G6 and G7 servers, you may indirectly set Power Regular to one of the modes that can degrade VMware ESX performance. When using HP Power Profile pre-sets, selecting the "Balanced Power and Performance" or the "Minimum Power Usage" setting will put Power Regulator in Dynamic Power Savings mode or Static Low Power mode, respectively. This influences VMware performance and workload balancing software.

Choosing the HP Power Profile "Maximum Performance" setting activates the Power Regulator Static High Performance mode which does not affect VMware performance, but also disables other power savings features unrelated to P-state control. The HP Power Profile includes a "Custom" setting that doesn't disable these additional power savings features. We recommend that you change the Power Regulator option to Static High Performance independently, which automatically puts the HP Power Profile setting into the Custom mode.

Using Intel Turbo Boost Technology® with HP Power Regulator

Turbo Boost Technology is a feature of Intel Xeon processors. The technology allows a processor to enter a performance state higher than the specified maximum frequency of the processor (also called "dynamic overclocking"). The processor will enter higher states only if there is enough thermal and power headroom and only if enabled by the power management software. HP Dynamic Power Savings mode and HP Static High Power mode take full advantage of Turbo Boost Technology, beginning with Intel processor-based ProLiant G6 servers. You can enable Turbo mode through RBSU.

Power consumption and system performance

HP engineers have extensively tested performance on systems operating in different Power Regulator modes. For example, Figures 2 and 3 display the power and performance results for both an HP ProLiant DL380 G7 server and an HP ProLiant DL385 G7 under different levels of application load while in the Static Low Power, Static High Performance, and Dynamic Power Savings modes.

Figure 2 shows Power Regulator mode comparisons under different CPU utilization (performance) loads using the two-processor DL380 G7 server with Quad-core Intel Xeon 5670 2.93-GHz processors. At loads under 70% of the maximum performance, Static Low Power mode uses up to 8% less power than Static High Performance mode to provide the same performance. Figure 2 also shows that Dynamic Power Savings mode (slow response) uses the entire CPU performance range like Static High mode, but provides the same power savings as Static Low mode.

Figure 2. Power and performance comparison of the DL380 G7 server in Dynamic Power Savings mode (slow and fast response time) and Static modes (Low Power and High Performance)

Dynamic Power Savings mode (Slow response) ——
Dynamic Power Savings mode (Fast response) ——
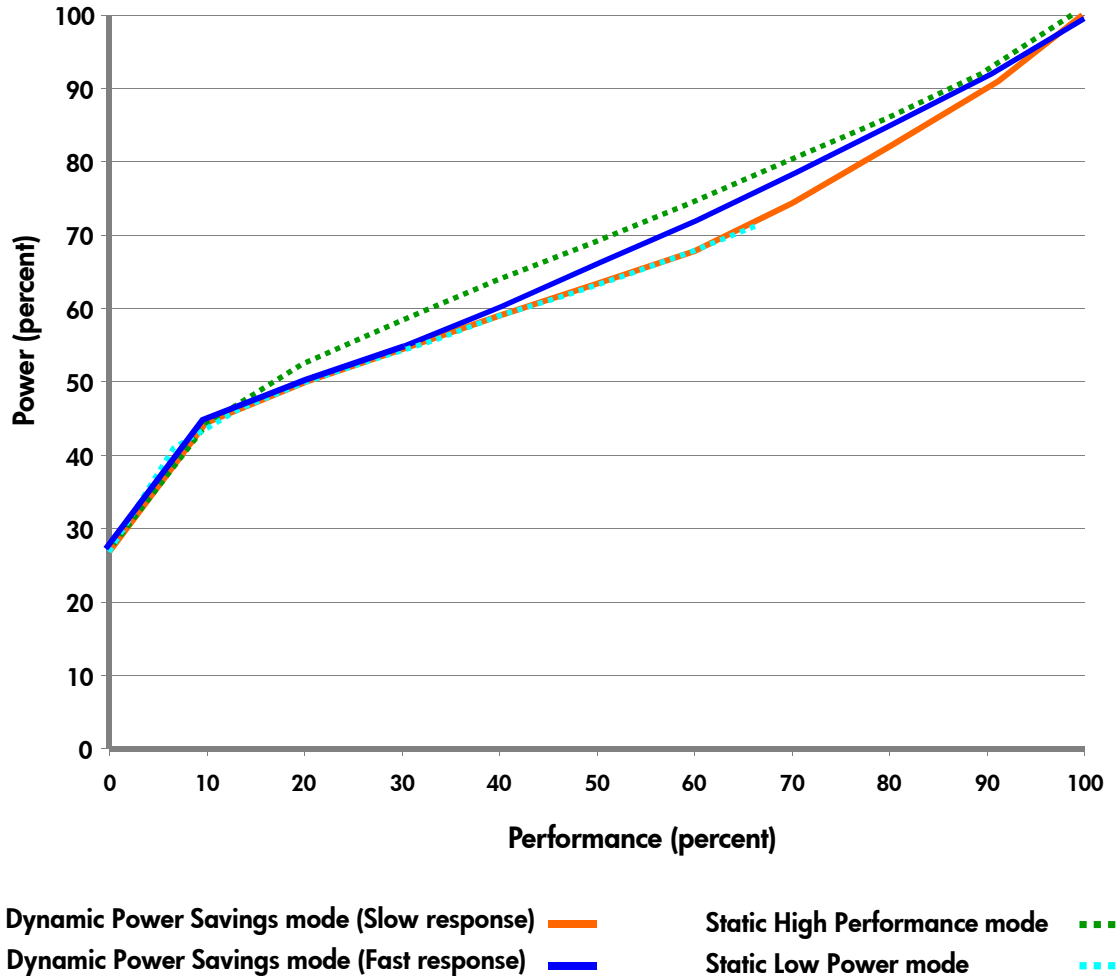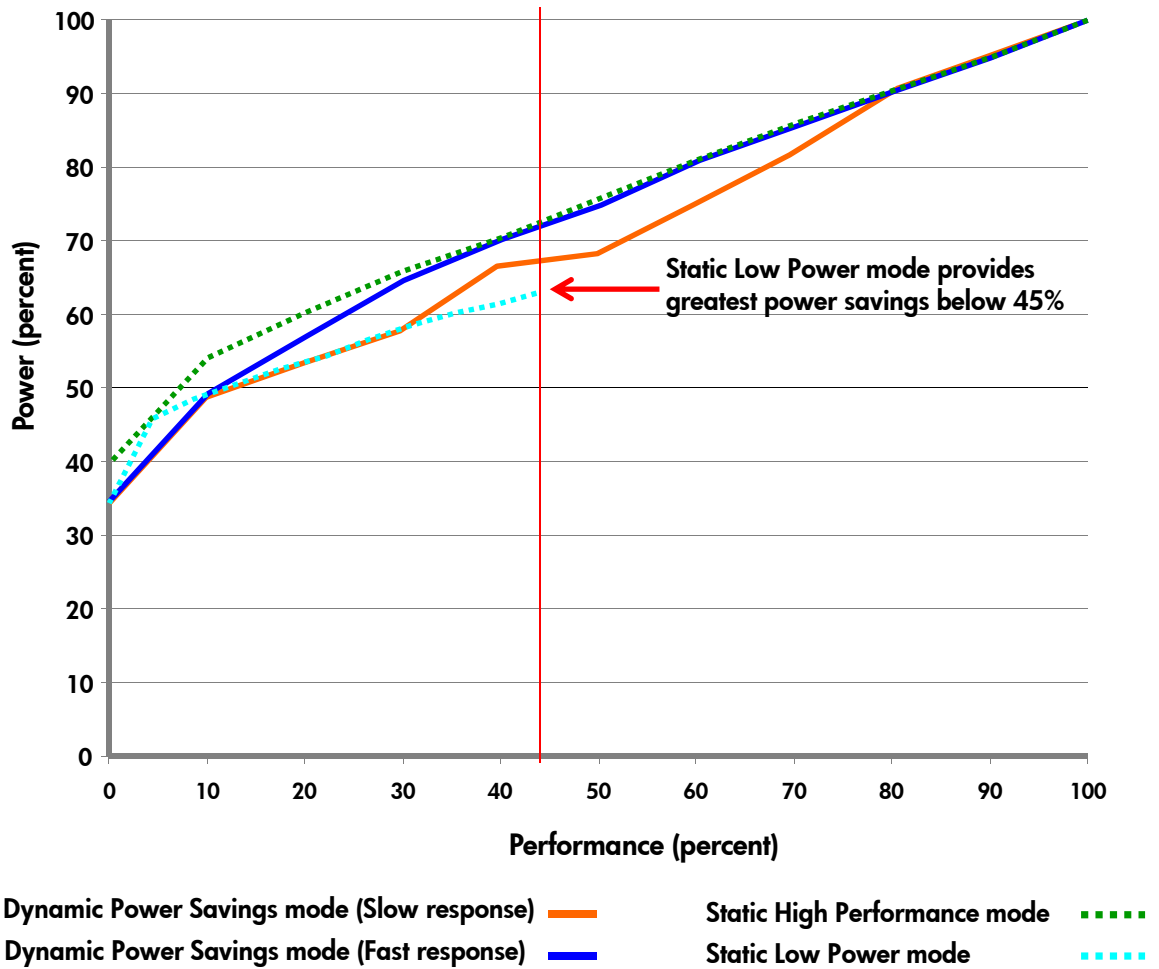Static High Performance mode ·····
Static Low Power mode ·····

Figure 3 shows the same power and performance comparison for the DL385 G7 server. In this comparison, if you don't require the full performance range of the processors, Static Low Power mode will give you the most power savings until you reach a performance utilization of about 45%. Above this level, the Dynamic Power Savings mode begins to run the processors at the higher P-states to maintain optimal system performance.

Figure 3. Power and performance comparison of the DL385 G7 server in Dynamic Power Savings mode (slow and fast response time) and Static modes (Low Power and High Performance)

You can draw several conclusions from these tests:

- Static High Performance mode delivers the highest performance at every performance interval up to maximum processor performance (100%), and it always consumes the most power.
- Static Low Power mode gives you the best power efficiency below a certain performance level (the level varies by processor type and maker). Static Low Power mode also gives you the best overall power efficiency if you don't require the full performance of the processor. The mode can provide significant power savings with little or no degradation in system performance at these lower utilization levels. Figures 2 and 3 show the processor performance limits for Static Low Power mode in each test.
- Because Dynamic Power Savings mode lets the firmware determine the best performance-to-power level, it achieves lower power consumption in low processor-utilization situations, and high performance when the processor reports high levels of utilization. These characteristics give

Dynamic Power Savings mode the most efficient overall power consumption and performance characteristics. The difference in processor performance (as measured by computational throughput) is negligible between the Static High mode and Dynamic Savings mode.

The application workload used in Figures 2 and 3 changed slowly enough over time that it favored the slow response time setting of Dynamic Power Saving mode. Dynamic Power Savings mode (fast response) is less power-efficient than slow response for application such as these but produces lower latency. Note that Dynamic Power Savings mode response times (both slow and fast) give you the ability to trade off levels of efficiency and latency relative to application workloads that may be more latency-sensitive.

Reducing power and cooling cost

Power Regulator is a valuable tool to improve server power use for typical client-server applications such as database, email, file and print, web server, and online transaction processing.
Figure 4 shows the effect of Power Regulator on a DL380 G7server running a workload comparable to the SPECpower_ssj2008 benchmark. The G7 system used in this test included two Quad-core Intel Xeon 5670 2.93-GHz processors with four 4 GB UDIMMs and a 60 GB SSD. The processors ran at 2.93 GHz in Static High Performance mode. The processor speed ranged from 1.6 to 2.93 GHz in Dynamic Power Savings mode.

Based on the test results shown in Figure 4, the power reduction for the DL380 G7 server configuration is 16 watts at a 60% application load.

Figure 4. Effective power consumption of Dynamic Power Savings and Static High Performance modes on a DL380 G7 server at various system loads
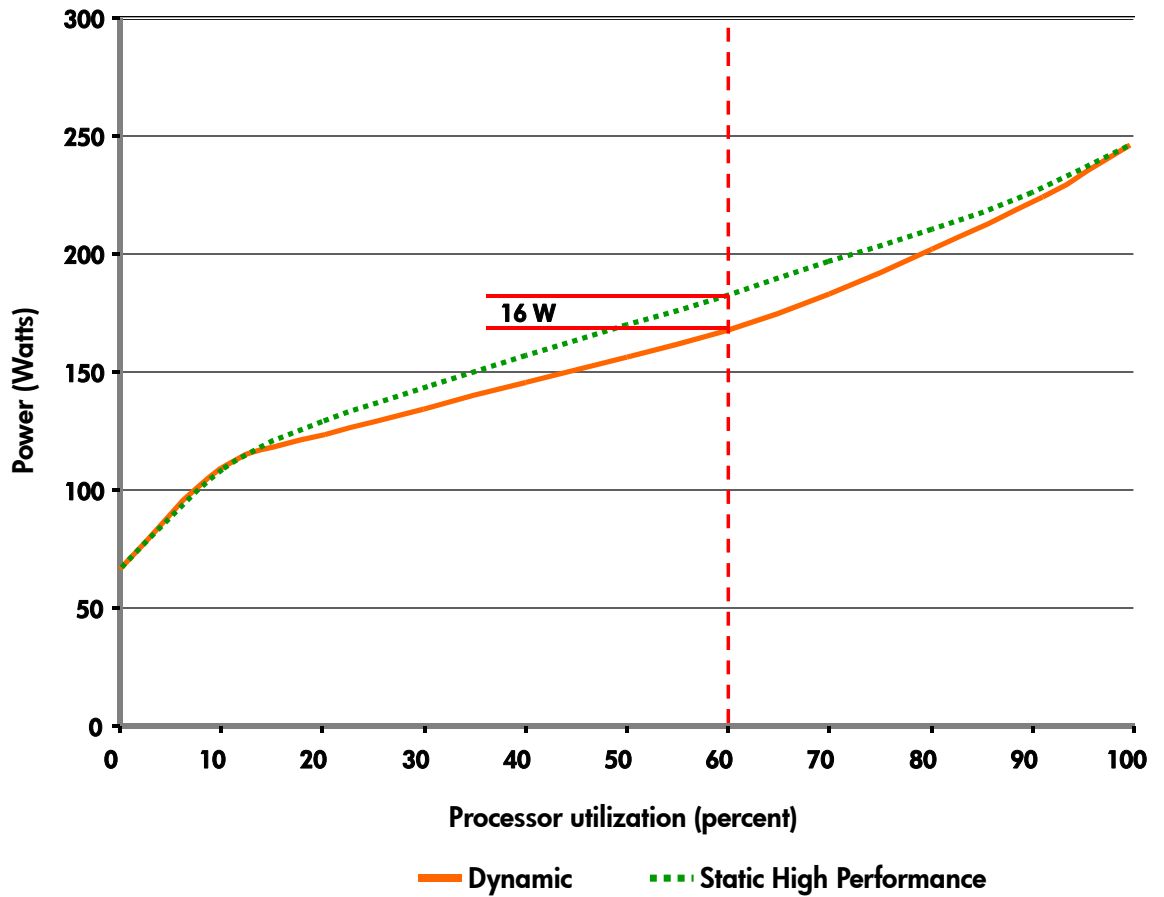

Table 3 shows the estimated cost savings of $28.03 per year on this system under 60% application loads in Dynamic Power Savings mode compared to Static High Performance mode. Actual savings vary depending on system configuration, processor utilization, data center power and cooling efficiency, and electricity cost. The cost savings should scale linearly with the number of servers and create significant savings for large data center operations.

Table 3. Calculated annual power cost savings for an HP ProLiant DL380 G7 server running the 60% average workload shown in Figure 4

Parameter	Calculation	Savings
Power saved		16 W
PUE* = 2.0	2.0 x 16 W	32 W (or 0.032 kW)
Annual energy saved	0.032 kW x 24 hrs/day x 365 days/yr	280.32 kWh/yr
Dollars saved (at 10 cents per kWh)	280.32 kWh/yr * $0.10)	$28.03/yr

* The Power Usage Effectiveness (PUE) metric describes data center power and cooling energy efficiency. A PUE of 2.0 is a typical level of energy efficiency in today's data centers.

Summary

HP Power Regulator for ProLiant servers is an important technology for any data center facing power and cooling challenges as well as any user interested in reducing power consumption and cost. Its flexible power saving modes and OS-independent ROM operation give you a broader tool set than OS-based power management. Its compatibility with other HP management products such as iLO and Insight Control makes the HP Power Regulator an integral part of a complete enterprise management solution.

For more information

Resource description	Web address
HP Insight Control power management User Guide	http://h20000.www2.hp.com/bc/docs/support/SupportManual/c02048495/c02048495.pdf
HP ProLiant G6 Server Power Management on Microsoft Windows Server 2008 R2	http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA2-5004ENW.pdf
Green Grid Data Center Power Efficiency Metrics: PUE and DCiE	http://thegreengrid.org/en/Global/Content/white-papers/The-Green-Grid-Data-Center-Power-Efficiency-Metrics-PUE-and-DCiE
Usage and Public Reporting Guidelines for the Green Grid Infrastructure Metrics (PUE/DCiE)	http://thegreengrid.org/en/Global/Content/white-papers/Usage%20and%20Public%20Reporting%20Guidelines%20for%20PUE%20DCiE
Implementing Microsoft Windows Server 2008 R2 on HP ProLiant servers, integration note	http://h20000.www2.hp.com/bc/docs/support/SupportManual/c01639594/c01639594.pdf
Processors supported by Power Regulator	www.hp.com/servers/power-regulator
SPECpower_ssj2008 benchmark	www.spec.org/power_ssj2008/

Call to action

Send comments about this paper to TechCom@HP.com

Follow us on Twitter: http://twitter.com/ISSGeekAtHP