Lenovo

The Importance of UEFI Settings and Hardware C-States

Last Update: 13 August 2017

Explains how to use C-states settings

Explains the relationships among power settings

Provides suggested system settings

Describes how to manage latency

Dan Woodruff

LENOVO PRESS

Abstract

To manage system behavior, an administrator requires an understanding of selectable UEFI and operating system settings. Many customers use default settings, yet the settings can be modified to optimize system operation. Implementing alternate settings must be done with care.

In this paper we briefly describe the functions that manage system behavior and how they interact with the system hardware. This paper is suitable for users and support personnel that have a basic understanding of system hardware architecture.

At Lenovo® Press, we bring together experts to produce technical publications around topics of importance to you, providing information and best practices for using Lenovo products and solutions to solve IT challenges.

See a list of our most recent publications at the Lenovo Press web site:

http://lenovopress.com

Do you have the latest version? We update our papers from time to time, so check whether you have the latest version of this document by clicking the **Check for Updates** button on the front page of the PDF. Pressing this button will take you to a web page that will tell you if you are reading the latest version of the document and give you a link to the latest if needed. While you're there, you can also sign up to get notified via email whenever we make an update.

Contents

Introduction

When designing systems, Lenovo engineers architect and document how to configure and tune Lenovo servers for energy efficiency, performance, and low latency using selective configuration settings. Many of these tuning parameters are accessible from UEFI in the System Setup menus.

One of these settings, Hardware C-states, is normally used to allow or disable manipulation of core frequency and voltage. It also can be used to quickly identify problem areas and mitigate the issue until a precise fix is provided.

This document, although focused on Lenovo high-performance servers, may also apply to systems produced by other manufacturers. Its intent is to:

1. Briefly describe how to manage system behavior related to C-states and latency.
2. At a high level, state what is affected via management settings.
3. Describe how to effectively use the C-states setting and when such use is beneficial.
4. Define general misuse of the C-states setting.
5. Provide some suggested system settings to resolve known issues and avoid using the C-states setting unnecessarily.

We recommend reviewing the reference material listed at the end of this document in conjunction with reading this paper.

Managing system behavior and latency

Latency is the time delay between when a system request is issued and when the system responds or starts execution of the action. Similarly, as a system takes time to go from a sleep state to fully awake, computer hardware requires time to change from one state to another. Figure 1 on page 4 depicts the tree for the various power groups for Intel CPU-based systems. Latency occurs when moving between the states within a group, if allowed by the set value for the group.

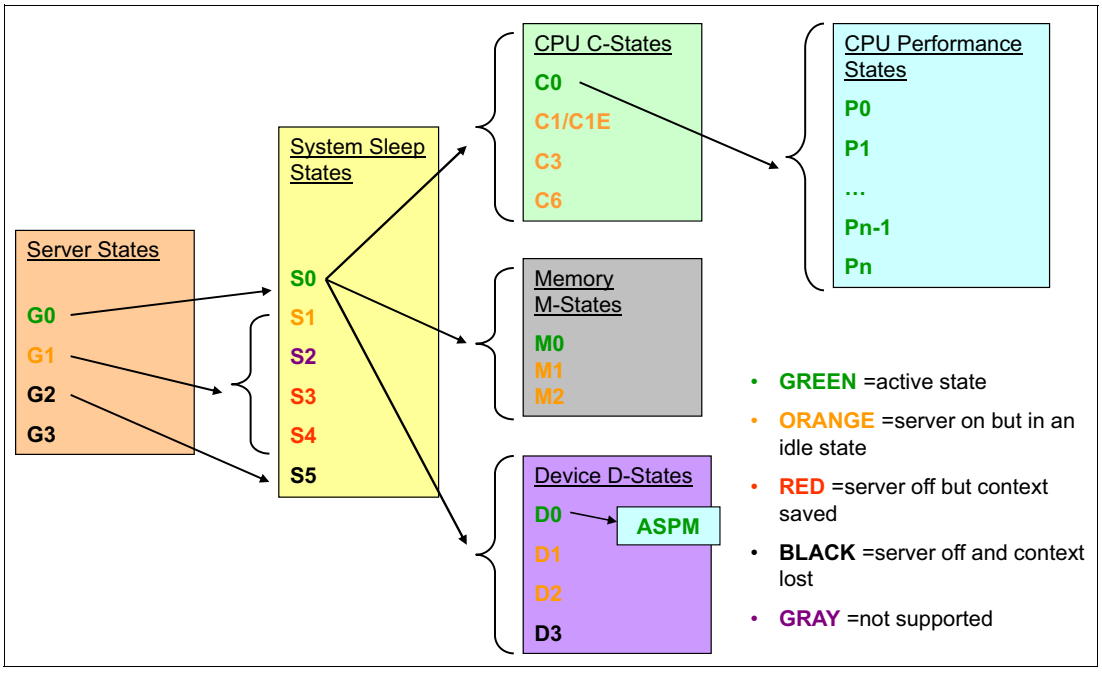

Figure 1 System states

As discussed in the section “Relationships among the power states” on page 29 of the Lenovo Press paper *Energy Efficiency Features of Lenovo System x Servers*, explains that the state a power group can attain may be dependent on the set value of another group and how latency changes.

https://lenovopress.com/redp5259-energy-efficiency-features-of-system-x-servers

There is a hierarchy among the power states. At the highest level, the G-states represent the overall state of the server. The G-states map to the S-states (system sleep states). Progressing to the right side of Figure 1, there are subsystem power states that represent the current state of the CPU, memory, and subsystem devices.

As shown by the arrows in Figure 1 above, certain power states cannot be entered if higher-level power states are inactive. For example, for a CPU core to be in P1 state, the CPU core also must be in C0 state, the system must be in S0 system sleep state, and the overall server must be in G0 state. With any of the power saving states, there is a trade-off between energy savings and latency.

For instance, enabling the CPU C6 state allows CPU cores to be turned off, which saves energy. However, because the CPU cores are powered down, it takes more time to restore their state when they transition back to the C0 state.

If maximum overall performance is desired, all power-saving states can be disabled. Although this configuration minimizes the latencies to transition into and out of the various power states, energy use is increased. At the other extreme, if power settings are optimized for maximum energy savings, performance can suffer because of long latencies.

For most applications, the default system settings offer a good balance between performance and efficiency. On most Lenovo servers, the defaults can be changed if either increased performance or power savings is preferred.

Modern operating systems now include the ability to monitor, manage and report system behavior. Depending on OS kernel boot parameters, some OSes can ignore set BIOS/UEFI

values (including C-states) and assign values according to installed application performance guidelines. They can also ensure adherence to latency limits and flag system issues when a system is unable to operate within those limits.

Important points:

- Research each UEFI setting and learn how the setting affects latency before changing the configuration
- New hardware and software technology releases can add or change C-states selections (for example; C-States: Legacy/Autonomous/Disable for Intel Xeon v4 (Broadwell) CPUs)
- Failure to understand the effects of the new values can result in unexpected system behavior, such as reboots, PSOD, BSOD, and hangs
- Newer generations of Lenovo servers may behave differently than previous generations with respect to C-states.
- Different generations of Lenovo servers will have different default settings.

What is affected by behavioral management

When power states change, multiple actions take place. For example:

- Power devices (e.g., voltage regulators) for processor packages start switching on and off to meet demand or limit the power to processor cores
- Memory might also switch between states, going in and out of self-refresh mode
- Other devices, such as adapters or storage, might switch between states as well

With the wide variety of optional hardware that can be configured as part of a system, signal timing in the code becomes critical. If any of these devices fail to behave according to their designed specifications, they can cause unexpected system behavior.

Use of the C-states setting

With Intel Xeon v4 processors, the Lenovo default Operating Mode Hardware C-States setting is Autonomous. The default setting for older processor types is: Enable. The setting provides management capabilities, but the setting has additional uses as well.

For example, some systems containing older processors encountered issues when the C-states setting was Enable. Changing the setting to Disable resolved those issues by eliminating CPU latency and removing hardware latency that may have contributed to the issue. The setting change provided a quick, temporary resolution for the issue under investigation. It narrowed the investigative focus and reduced the time needed to find the root cause of the issue. When the final fix was released and implemented, the C-states setting was able to be set back to the original Enable value.

The C-states setting can be used in the following ways:

- To provide CPU power management capabilities
- Set to Disable, it can aid in determining where an issue exists.
- Set to Disable, it can be used to temporarily mitigate some issues.

The advantages of using the setting this way are that it offers a short-term benefit and latency is minimized. The disadvantage of using the setting this way is that system behavior may not be what is wanted (for example, energy consumption is increased).

Misuse of the C-states setting

A review of support cases indicates that administrators sometimes change the C-States setting from the default, based on their experience. With newer systems this may not always be appropriate. It is important to understand the behavior controlled by each setting and how it would affect system operations. Lenovo recommends consulting system documentation, Lenovo Tech Tips, IBM RETAIN tips, or knowledgeable people prior to changing the value.

Always setting C-states to Disabled to proactively avoid issues (that may not exist), is unnecessary, possibly counterproductive, and circumvents the purpose of having the selection.

Suggested UEFI settings for operating systems

Although the intent of this document is to point out how the C-states setting should be used, other BIOS/UEFI settings have multiple uses as well. They, too, can be used to mitigate known issues.

The following UEFI suggested settings can be made via the F1 settings menu. For more information about these settings, see the product publications that come with the server.

Windows

Suggested UEFI settings via the F1 menu (Figure 2 on page 6):

/System Settings/Recovery and RAS/Advanced RAS/Machine Check Recovery = **Disable**

/System Settings/Recovery and RAS/Advanced RAS/PCI Live Error Recovery = **Disable**

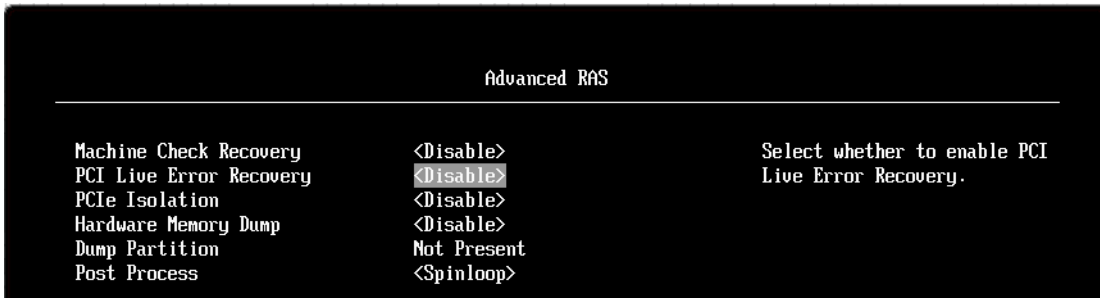

Figure 2 Windows-specific settings

Setting both of these to **Disable** has been proven to resolve numerous Windows blue screen stop errors (BSODs), hangs, and unexpected reboots. Investigation into why disabling the settings resolves some issues has shown it to be related to failure by the user to install critical updates or failure to install the updates in the proper order.

VMware ESXi

Suggested UEFI settings via the F1 menu (Figure 3):

/System Settings/Devices and I/O Ports/PCI 64-Bit Resource Allocation = **Disabled**

/System Settings/Devices and I/O Ports/MM Config Base = **3G**

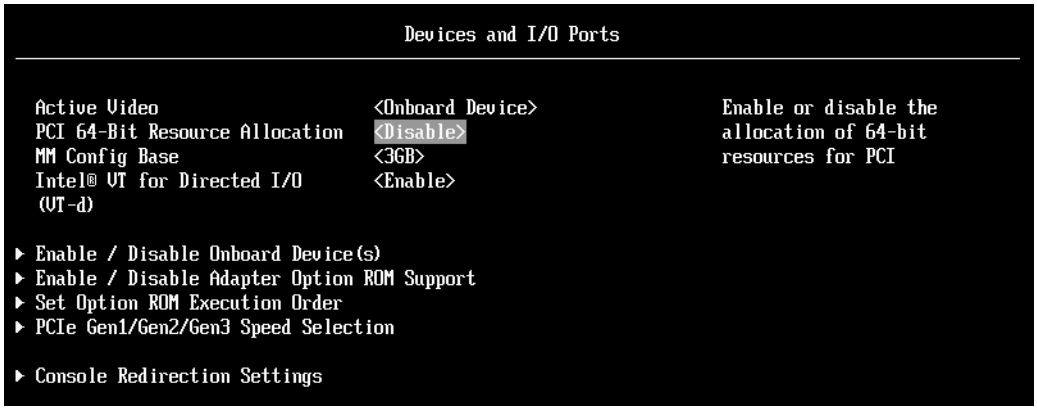

Figure 3 VMware-specific settings

The following link instructs installers to set the above settings for a known ESXi 6.0 U1 installation problem:

https://support.lenovo.com/us/en/solutions/HT118267

Lenovo VMware Support Specialists also strongly recommend:

- Use the same settings for VMware ESXi versions 5.5 and higher.
- Apply the settings for normal runtime and not just installation, to prevent stop errors (PSODs), PCI related issues, hangs, and unexpected reboots.

Linux

To prevent the OS from ignoring the BIOS/UEFI settings:

For Red Hat Enterprise Linux (RHEL):

- Edit /boot/grub/grub.conf and add the **`intel_idle.max_cstate=0`** parameter to the kernel line.

For SUSE Linux Enterprise Server (SLES):

- Edit /boot/grub/menu.lst and add the **`intel_idle.max_cstate=0`** parameter to the kernel line.

References

- Energy Efficiency Features of Lenovo System x® Servers

  https://lenovopress.com/redp5259
- Lenovo System x3850 X6 and x3950 X6 Planning and Implementation Guide

  https://lenovopress.com/sg248208-x6-planning-and-implementation-guide

- Tuning IBM System x Servers for Performance

  https://lenovopress.com/sg245287

Change history

August 13, 2017:

- Minor corrections

Author

Dan Woodruff is a team lead in product engineering for Lenovo high performance X6 servers. He is responsible for Level 3 support of the servers to resolve all complex issues that customers encounter. He has been in this role in Lenovo and IBM since 2004. Prior to this role, he was the product engineering lead for the development and introduction of the IBM iDataPlex® product family, and an engineer for IBM World-Wide Manufacturing Test Engineering, developing test processes and test code for IBM x86 servers. Dan has worked in the IT industry since 1980. He has an Associates Degree in Electrical Engineering (ASEE) from Muskingum Techinical College (1978) and a Bachelor of Science in Electrical Engineering from Ohio University-Athens (1980).

Thanks to the following people for their contributions to this project:

- Mark Chapman
- David Watts, Lenovo Press

Notices

Lenovo may not offer the products, services, or features discussed in this document in all countries. Consult your local Lenovo representative for information on the products and services currently available in your area. Any reference to a Lenovo product, program, or service is not intended to state or imply that only that Lenovo product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any Lenovo intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any other product, program, or service.

Lenovo may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

Lenovo (United States), Inc.
1009 Think Place - Building One
Morrisville, NC 27560
U.S.A.
Attention: Lenovo Director of Licensing

LENOVO PROVIDES THIS PUBLICATION “AS IS” WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some jurisdictions do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. Lenovo may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

The products described in this document are not intended for use in implantation or other life support applications where malfunction may result in injury or death to persons. The information contained in this document does not affect or change Lenovo product specifications or warranties. Nothing in this document shall operate as an express or implied license or indemnity under the intellectual property rights of Lenovo or third parties. All information contained in this document was obtained in specific environments and is presented as an illustration. The result obtained in other operating environments may vary.

Lenovo may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Any references in this publication to non-Lenovo Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this Lenovo product, and use of those Web sites is at your own risk.

Any performance data contained herein was determined in a controlled environment. Therefore, the result obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurements may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

This document was created or updated on August 13, 2017.

Send us your comments via the **Rate & Provide Feedback** form found at
http://lenovopress.com/lp1234

Trademarks