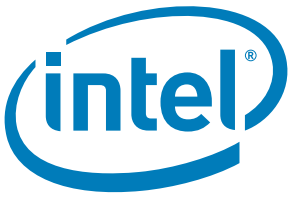

Timestamp-Counter Scaling for Virtualization White Paper

This document is intended only for VMM or hypervisor software developers and not for application developers or end-customers. Readers are expected to be knowledgeable about Intel® Architecture and Intel® Virtualization Technology.

Reference Number: 333159-001

September 2015

CONTENTS

CHAPTER 1
INTRODUCTION

1.1 OVERVIEW

This paper describes an Intel® Virtualization Technology (Intel® VT) enhancement for future Intel processors. This feature, referred to as **timestamp-counter scaling** (**TSC scaling**), further extends the capability of virtual-machine monitor (VMM) software that employs the TSC-offsetting mechanism by allowing that software finer control over the value of the timestamp counter (TSC) read during guest virtual machine (VM) execution. Details of Intel VT, including TSC offsetting, can be found in *Intel® 64 and IA-32 Architectures Software Developer's Manual* (SDM) in Volume 3C.

Timestamp-counter offsetting (**TSC offsetting**) is an existing feature that allows VMM software to specify a value (the **TSC offset**) that is added to the TSC when it is read by guest software. A VMM can use this feature to provide guest software with the illusion that it is operating at a time later or earlier than that represented by the current TSC value.

With TSC offsetting, guest software perceives a TSC that is offset from the real hardware, but which advances at the same rate. That may be adequate for usages in which the offset is used to account for execution time before virtual machine was created. But it might not suffice if the VMM migrates a virtual machine between platforms on which the TSC moves at different rates.

TSC scaling provides VMM software with a mechanism by which is it can adjust the TSC rate perceived by guest software. When TSC scaling and TSC offsetting are both enabled, reads from the TSC in VMX non-root operation multiply the actual TSC value by a new **TSC multiplier**, add the TSC offset to the product, and return the sum to guest software.

With both TSC offsetting and TSC scaling, a VMM that migrates a virtual machine from one platform to another can configure the TSC offset and the TSC multiplier on the new platform so that the TSC (as perceived by the guest) appears to proceed from the same value that it had before the migration **and at the same rate**.

The remainder of this paper is organized in subsections which cover specific changes in VMX to support TSC scaling:

- Section 1.2 details changes to the VMCS introduced with the TSC-scaling feature.
- Section 1.3 details changes to VMX non-root operation.
- Section 1.4 describes changes to VM entries (there are no changes to VM exits).
- Section 1.5 details changes to the capability reporting introduced with TSC-scaling.

1.2 VMCS CHANGES

Secondary processor-based VM-execution control 25 is defined as **TSC scaling**.

A new 64-bit VM-execution control field called the **TSC multiplier** is defined. The VMCS-field encoding pair for the TSC multiplier is 00002032H (for all 64 bits) and 00002033H (for the upper 32 bits).

The TSC multiplier field exists only on processors that support the 1-setting of the "TSC scaling" VM-execution control.

1.3 CHANGES TO VMX NON-ROOT OPERATION

If the "TSC scaling" VM-execution control is 1, the behavior in VMX non-root operation is modified as described in this section.[1]

The modifications apply only if the "TSC offsetting" VM-execution control is 1. When the modifications apply, they affect instructions that normally return in EDX:EAX the value of the timestamp counter. These instructions are RDMSR (ECX contains 10H, indicating the IA32_TIME_STAMP_COUNTER MSR), RDTSC, and RDTSCP.[2]

Specifically, an affected read from the timestamp counter first computes the product of the value of the IA32_TIME_STAMP_COUNTER MSR and the value of the TSC multiplier. It then shifts the value of the product right by 48 bits and loads EAX:EDX with the sum of that shifted value and the value of the TSC offset.

TSC scaling applies also to any timestamp-counter values that the processor may record in records produced for precise event-based sampling (PEBS) and for Intel Processor Trace (PT). Any timestamp-counter value that one of these records would contain (if produced in VMX non-root operation) is first multiplied by the TSC multiplier and then added to the TSC offset, as described above.

1.4 CHANGES TO VM ENTRIES

If the "activate secondary controls" and "TSC scaling" VM-execution controls are both 1, VM entries ensure the TSC multiplier is not zero.

VM entry fails if this check fails. When such a failure occurs, control passes to the next instruction, RFLAGS.ZF is set to 1 to indicate the failure, and the VM-instruction error field is loaded with value 7, indicating "VM entry with invalid control field(s)."

This check may be performed in any order with respect to other checks on VMX controls and the host-state area. Different processors may thus give different error numbers for the same VMCS.

The "TSC scaling" VM-execution control may be 1 even if the "TSC offsetting" VM-execution control is 0; VM entry will not fail because of this condition.

1.5 CHANGES TO VMX CAPABILITY REPORTING

Section 1.3 specified that bit 25 of the secondary processor-based VM-execution controls is defined as "TSC scaling". A processor that supports the 1-setting of "TSC scaling" sets bit 57 of the IA32_VMX_PROCBASED_CTLS2 MSR (index 48BH). RDMSR of that MSR returns 1 in bit 25 of EDX.

1. "TSC scaling" is a secondary processor-based VM-execution control. If bit 31 of the primary processor-based VM-execution controls ("activate secondary controls") is 0, VMX non-root operation functions as if all secondary processor-based VM-execution controls (including "TSC scaling") were 0.
2. The modifications apply to RDTSC and RDTSCP only if the "RDTSC exiting" VM-execution control is 0.