

A Performance Comparison Between Enlightenment and Emulation in Microsoft Hyper-V

Hasan Fayyad-Kazan¹, Luc Perneel² and Martin Timmerman³

¹ Vrije Universiteit Brussel

Received: 10 December 2012 Accepted: 3 January 2013 Published: 15 January 2013

Abstract

Microsoft (MS) Hyper-V is a native hypervisor that enables platform virtualization on x86-64 systems. It is a micro-kernelized hypervisor where a host operating system provides the drivers for the hardware. This approach leverages MS Hyper-V to support enlightenments (the Microsoft name for Paravirtualization) in addition to the hardware emulation virtualization technique. This paper provides a quantitative performance comparison, using different tests and scenarios, between enlightened and emulated Virtual Machines (VMs) hosted by MS Hyper-V server 2012. The experimental results show that MS enlightenments improve performance by a factor of more than two.

Index terms— virtualization, hyper-v, enlightenments, emulation.

1 Introduction

Virtualization has become a popular way to make more efficient use of server resources within both private data centers and public cloud platforms. It refers to the creation of a Virtual Machine (VM) which acts as a real computer with an operating system (OS) [1]. It also allows sharing the underlying physical machine resources with different VMs.

The software layer providing the virtualization is called a Virtual Machine Monitor (VMM) or hypervisor [1]. It can be either Type 1 (or native, bare metal) running directly on the host's hardware to control the hardware and to manage guest operating systems, or Type 2 (or hosted) running within a conventional operating-system environment.

Since it has direct access to the hardware resources rather than going through an operating system, a native hypervisor is more efficient than a hosted architecture and delivers greater scalability, robustness and performance [2].

Microsoft Hyper-V implements Type 1 hypervisor virtualization [3]. In this approach, a hypervisor runs directly on the hardware of the host system and is responsible for sharing the physical hardware resources with multiple virtual machines [4]. In basic terms, the primary purpose of the hypervisor is to manage the physical CPU(s) and memory allocation between the various virtual machines running on the host system.

There are several ways to implement virtualization. Two leading approaches are Full virtualization (FV)/Hardware emulation and Para-virtualization (PV) [5]. Enlightenment is the Microsoft name for Paravirtualization.

This paper provides a quantitative performance comparison between hardware emulation and Enlightenment (Para-Virtualization) techniques hosted by MS Hyper-V server 2012.

It is organized as follows: Section 2 describes MS Hyper-V architecture and Enlightenment approach; Section 3 shows the experimental setup used for our evaluation; Section 4 explains the test metrics, scenarios and results obtained; and section 5 gives a final conclusion.

II.

3 Microsoft Hyper-v

Microsoft Hyper-V is a hypervisor-based virtualization technology for x64 versions of Windows Server [6]. It exists in two variants: as a stand-alone product called Hyper-V Server and as an installable role/component in Windows Server [7].

There is no difference between MS Hyper-V in each of these two variants. The hypervisor is the same regardless of the installed edition [7].

MS Hyper-V requires a processor with hardware-assisted virtualization functionality, enabling a much more compact virtualization codebase and associated performance improvements [3].

The Hyper-V architecture is based on microkernelized hypervisors (figure 1). This is an approach where a host operating system, referred to as the parent partition, provides management features and the drivers for the hardware [8]. (D D D D D D D)

Year Figure 1 : Hyper-V Architecture [6] With this approach, the only layer between a guest operating system and the hardware is a streamlined hypervisor with simple partitioning functionality. The hypervisor has no third-party device drivers [9]. The drivers required for hardware sharing reside in the host operating system, which provides access to the rich set of drivers already built for Windows [9].

MS Hyper-V implements isolation of virtual machines in terms of a partition (operating system and applications). A hypervisor instance has to have at least one parent partition, running a supported version of Windows Server [6]. The virtualization stack runs in the parent partition and has direct access to the hardware devices. The parent partition then creates the child partitions which host the guest OSs [6].

Child partitions do not have direct access to hardware resources. Hyper-V can host two categories of operating systems in the child partitions: Enlightened (Hyper-V Aware) and un-enlightened (Hyper-V Unaware) operating systems [10]. Enlightened partition has a virtual view of the resources, in terms of virtual devices. Any request to the virtual devices is redirected via the VMBus (figure 1) -a logical channel which enables inter-partition communication -to the devices in the parent partition managing the requests. Parent partitions run a Virtualization Service Provider (VSP), which connects to the VMBus and handles device access requests from child partitions [6]. Enlightened child partition virtual devices internally run a Virtualization Service Client (VSC) (figure1), which redirect the request to VSPs in the parent partition via the VMBus [6]. The VSCs are the drivers of the virtual machine, which together with other integration components are referred to as Enlightenments that provide advanced features and performance for a virtual machine. In contrast, the unenlightened child partition does not have the integration components and the VSCs; everything is emulated.

III.

5 Experimental Setup

Microsoft Hyper-V Server 2012 is tested here. It is a dedicated stand-alone product that contains the hypervisor, Windows Server driver model, virtualization capabilities, and supporting components such as failover clustering, but does not contain the robust set of features and roles found in the Windows Server operating system [11].

As MS Hyper-V supports enlightened and emulated VMs, both VMs are created, running Linux PREEMPT_RT v3.8.4-rt2 [12]. Being open source and configurable for usage in enlightened VM are the main reasons for selecting it as the guest OS. This also permits us to compare with another ongoing study of XEN.

Both Linux versions (for enlightened and emulated VM) are built using the buildroot [13] tool to make sure that the enlightenment drivers are added to the enlightened VM.

6 Global Journal of Computer Science and Technology

Volume XIII Issue II Version I The tests are done in each VM separately. Under Test VM (UTVM) is the name used for the tested VM, which can be either enlightened or emulated. Each VM has one virtual CPU (vCPU). machine (Bare-Machine) as a reference, using the same OS of the UTVM.

7 i. Clock tick processing duration

The kernel clock tick processing duration is examined here. The results of this test are extremely important as the clock interrupt -being on a high level interrupt on the used hardware platform -will bias all other performed measurements. Using a tickless kernel will not prevent this from happening as it will only lower the number of occurrences. The kernel is not using the tickless timer option.

Here is a description of how this test is performed: a real-time thread with the highest priority is created. This thread does a finite loop of the following tasks: starting the measurement by reading the time using RDTSC instruction, executing a "busy loop" that does some calculations and stopping the measurement by reading the time again using the same instruction. Having the time before and after the "busy loop" provides the time needed to finish its job. In case we run this test on the bare-machine, this "busy loop" will be delayed only by interrupt handlers. As we remove all other interrupt sources, only the clock tick timer interrupt can delay the "busy loop". When the "busy loop" is interrupted, its execution time increases.

Running the same test in a VM also shows when it is scheduled away by the VMM, which in turn impacts latency.

Figure 2 presents the results of this test on the baremachine, followed by an explanation. The X-axis indicates the time when a measurement sample is taken with reference to the start of the test. The Y-axis indicates the duration of the measured event; in this case the total duration of the "busy loop". The lower values (68 s) of figure 2 present the "busy loop" execution durations if no clock tick happens. In case of clock tick interruption, its execution is delayed until the clock interrupt is handled, which is 76 s (top values). The difference between the two values is the delay spent handling the tick (executing the handler), which is 8 s.

Note that the kernel clock is configured to run at 1000 Hz, which corresponds to a tick each 1 ms. This is obvious in figure 2, which is a zoomed version of figure 3 below. This test is very useful as it detects all the delays that may occur in a system during runtime. Therefore, we execute this test for long duration (more than one hour) to capture 50 million samples. The results in the tables of section D are comparing the maximum results obtained from the 50 million samples. The hardware platform used for conducting the tests has the following characteristics: Intel® Desktop Board DH77KC, Intel® Xeon® Processor E3-1220v2 with 4 cores each running at a frequency of 3.1 GHz, and no hyper-threading support. The cache memory size is as follows: each core has 32 KB of L1 data cache, 32KB of L1 instruction cache and 256 KB of L2 cache. L3 cache is 8MB accessible to all cores. The system memory is 8 GB. Below is an explanation of the evaluation tests. Note that the tests are initially done on a non-virtualized One physical CPU is allocated for each VM, using the "virtual machine reserve" and "virtual machine limit" attributes in the VM settings using Hyper-V Manager.

A Figure ?? : Thread switch latency between 2 threads on the Bare-machine Figure ?? shows that the minimum switch latency between 2 threads is around 0.43 s; the maximum latency is 11.45 s which is dependent on the clock tick processing duration.

Table ?? below shows the results of performing this test on the bare-machine using 2 and 1000 threads. Armstrong, Hyper-V Program Manager, explains in the blog "Processor Affinity and why you do not need it on Hyper-V" [14] that there is no need for this concept in Hyper-V. Instead, one can reserve a physical CPU (pCPU) for the VM to guarantee that it always has a whole processor.

Moreover, if a VM has one vCPU and the host has more than one core, this VM can be mapped to any of the available cores in a round-robin way between all the cores [15]. The parent partition is the only VM parked on core 0.

8 d) Testing scenarios

Below is a description of the scenarios used for the evaluation. In all the scenarios drawings, the parent partition (VM) is not shown because it is idle. As shown in figure 5, this scenario has only one VM, the UTMV with one vCPU. This vCPU can run on any core (physical CPU) during runtime. The aim of this scenario is to detect the pure hypervisor overhead (as there is no contention).

ii. Thread switch latency between threads of same priority This test measures the time needed to switch between threads having the same priority. Although realtime threads should be on different priority levels to be capable of applying rate monotonic scheduling theory [16], this test is executed with threads on the same priority level in order to easily measure thread switch latency without interference of something else.

For this test, threads must voluntarily yield the processor for other threads, so the SCHED_FIFO scheduling policy is used. If we didn't use the FIFO policy, a round-robin clock event could occur between the yield and the trace, and then the thread activation would not be seen in the test trace. The test looks for worst-case behavior and therefore it is done with an increasing number of threads, starting with 2 and going up to 1000. As we increase the number of active threads, the caching effect becomes visible as the thread context will no longer be able to reside in the cache.

Table ?? : Comparison between the "Thread switch latency" test results Note that the "maximum switch latency" in all the scenarios depends on the processing durations of clock tick and other interrupts that may occur in the system during the testing time.

ii. Scenario2: One-to-One As mentioned before, Hyper-V does not support affinity. It sends the workload of a VM to the first physical CPU that is available.

In this scenario, there is only one physical CPU available, while the other three are disabled from the BIOS. There is only the UTMV, together with the parent partition which is always parked on CPU-0 but idle. Therefore, UTMV is also pinned to CPU-0. The aim of this scenario is to clarify if the affinity technique removes the periodic high measurements. Figure ?? shows that every second, the hypervisor is doing some tasks/scheduling decisions which causes the VM to be suspended/scheduled-away resulting in such high values periodically.

Note that our policy is black-box testing which makes it difficult to understand the internal behavior of an out-of-the-box product.

The emulated VM behaves exactly the same except with higher values. Table 2 is a comparison between the "clock tick processing duration" test results for both VMs, while table 3 is a comparison for the test results. This scenario is exactly the same as scenario 3 except using Memory-Load VM instead of CPU-Load VM. This VM is running an infinite loop of memcpy() function that copies 9 MB (a value that is larger than the whole caches) from one object to another. The other goal of this scenario using such a VM is to detect the cache effects on

the performance the UTM. 8 and 9 compare the results of the two tests: "Clock Tick processing duration" and "Thread switch latency".

9 Table 8 : Comparison between the "clock-tick processing duration" test results

Table ?? shows that measurements of this scenario are greater than the ones of the previous scenario (scenario 3) by almost 3 ms. 6 and 7 compares the results of the two tests: "Clock Tick processing duration" and "Thread switch latency". The resulting values of this scenario are around three times greater than the ones of the previous scenario (scenario 5) even though number of VMs is running. This difference in the results is due to the concept explained in the following section (System bus bottleneck in SMP systems).

System bus bottleneck in SMP systems.

The hardware platform used for this evaluation is a Symmetric Multiprocessing (SMP) system with four identical processors connected to a single shared main memory using a system bus. They have full access to all I/O devices and are treated equally.

The system memory bus or system bus can be used by only one core at a time. If two processors are executing tasks that need to use the system bus at the same time, then one of them will use the bus while the other will be blocked for some time. As the processor used has 4 cores, when all of these are running at the same time, system bus contention occurs. Scenario 5 is not causing high overheads because the CPU stress program is quite small and fits © 2013 Global Journals Inc. (US)

10 Global Journal of Computer Science and Technology

Volume XIII Issue II Version I 27 () in the core cache together with its data. Therefore, the are not intensively loading the system bus which in turn will not highly affect the UTM.

Referring back to scenario 6, the three Memory-Load VMs are intensively using the system bus. The UTM is also running and requires the usage of system bus from time to time. Therefore, the system bus is shared most of the time between four VMs (UTVM and 3 Memory-Load VMs), which causes extra contention. Thus, the more cores in the system that are accessing the system bus simultaneously, the more contention will occur and thus the overhead increases.

To explicitly show this effect, we created another additional scenario (scenario 7 below) where only one Memory-Load VM is sharing the resources with the UTM. The following demonstrates our observation. vii. Scenario7: TWO-to-ALL with 1 Memory-Load VM This work compares the performance between the two types of VM. For this purpose, different tests and several scenarios are used. The results show that the enlightened VM performs on average twice as good as the hardware-emulated VM. This performance enhancement may increase/decrease depending on the scenario in question.

Even with VM performance is low compared with bare-machine (nonvirtualized) performance.

A shared-memory symmetric multiprocessor hardware with four physical cores is used for conducting the tests. The results also show that the synchronous usage of all the available cores causes an intensive overload in the system bus which in turn increases latencies by a factor of 3 when compared with a system with only one active core.

11 Conclusion

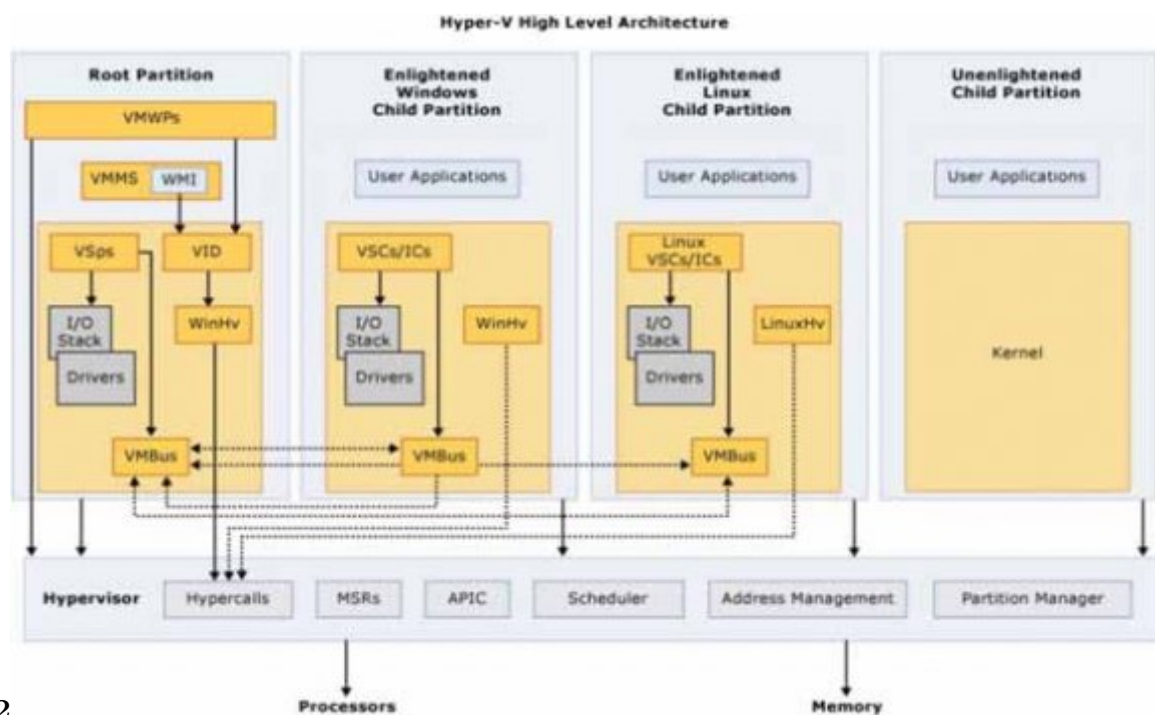
Hyper-V is a "Microkernelized Type 1" hypervisor which leverages paravirtualization (called Enlightenment by Microsoft) in addition to the traditional hardware emulation technique. It exists in two variants: as a stand-alone product called Hyper-V Server and as an installable role in Windows Server.

There is no difference between MS Hyper-V in each of these two variants. The hypervisor is the same regardless of the installed edition.

In



Figure 1: A



2

Figure 2: Figure 2 :

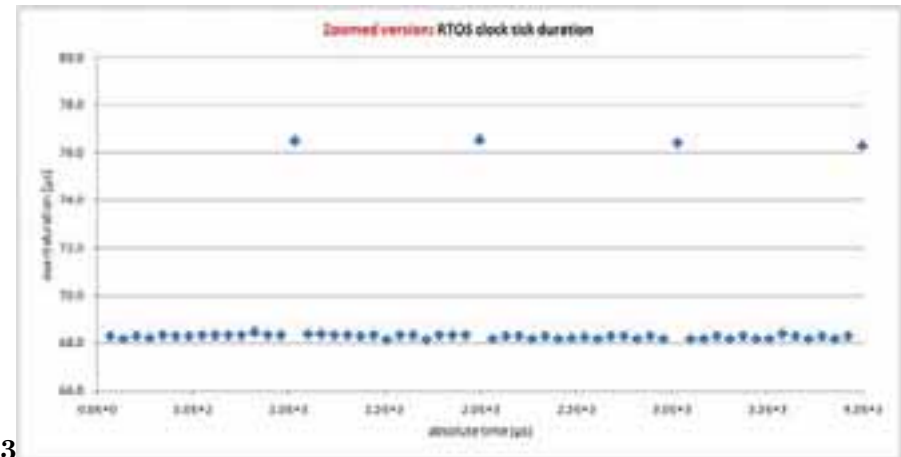


Figure 3: Figure 3 :

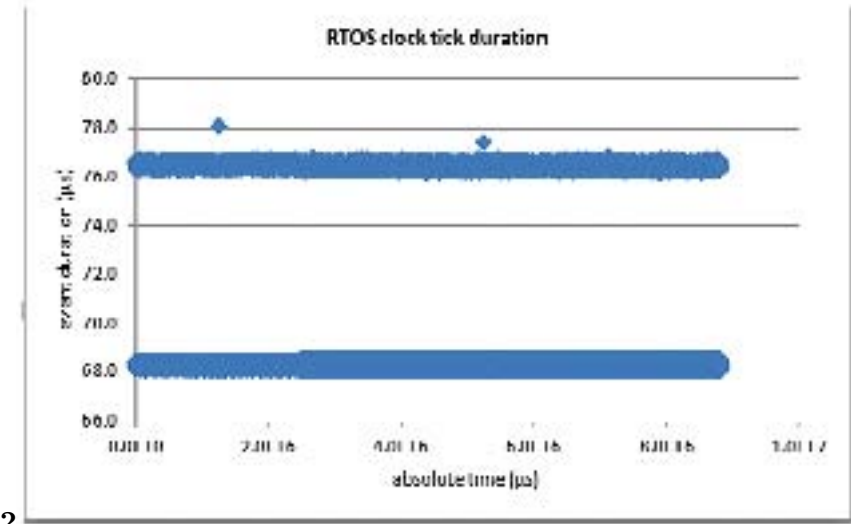


Figure 4: () 2 A

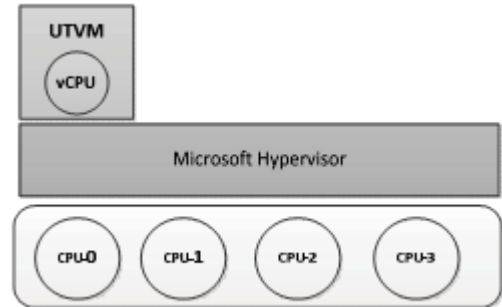


Figure 5:

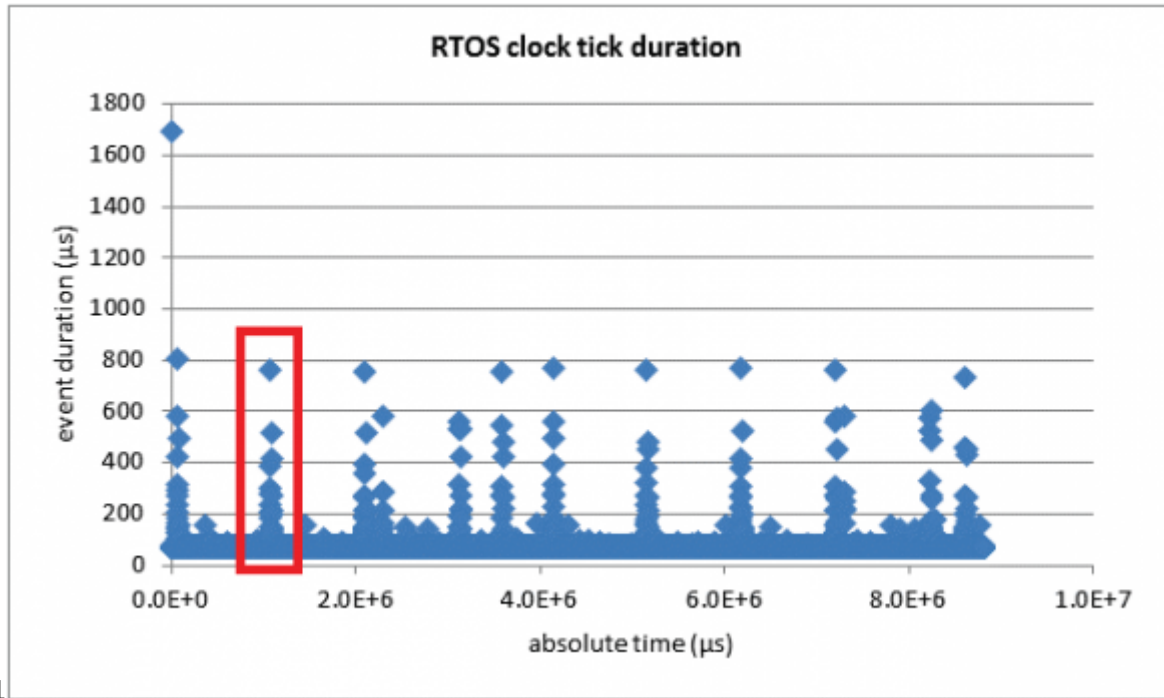


Figure 6: Table 1 :

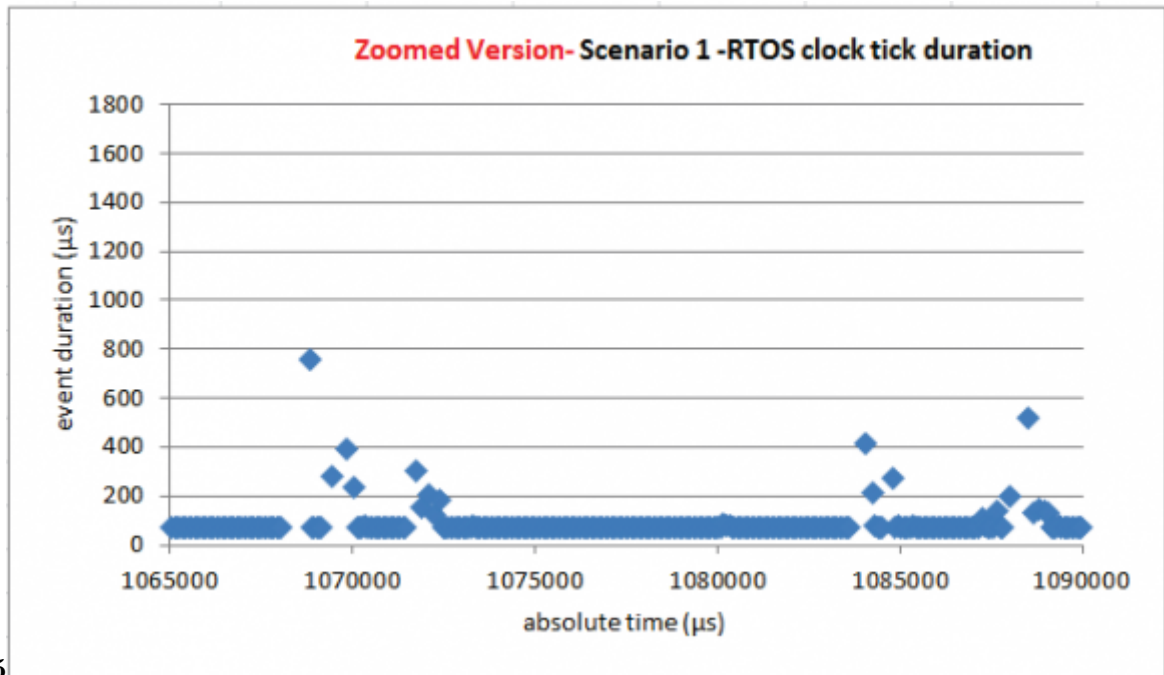


Figure 7: Figure 5 :

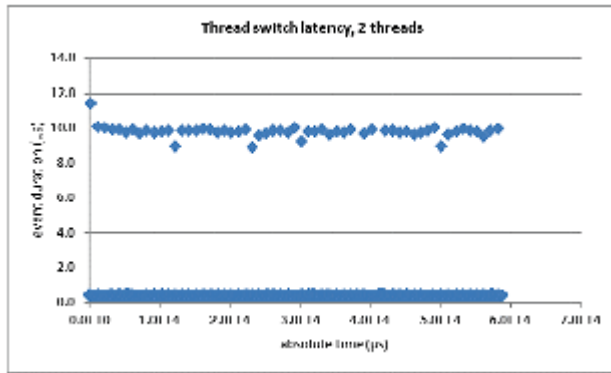


Figure 8: Figure 6 : 1 Figure 7 :A

Maximum Switch Latency between:		
	2 Threads	1000 Threads
Bare-Machine	11.2 μ s	11.45 μ s
Emulated VM	3.53 ms	1.24 ms
Enlightened VM	63 μ s	687 μ s

Figure 9: Figure 8 :

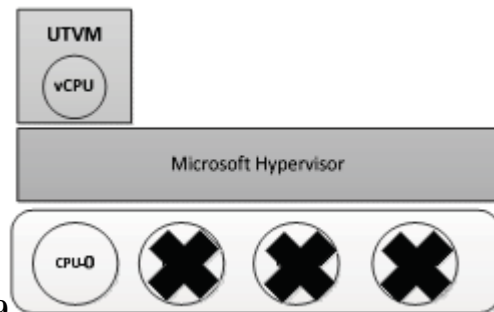


Figure 10: Figure 9 :)

Clock Tick Processing Duration	
	Maximum Overhead
Bare-Machine	10 μ s
Emulated VM	7.19 ms
Enlightened VM	4.08 ms

Figure 11: Figure 10 :

Maximum Switch Latency between:		
	2 Threads	1000 Threads
Bare-Machine	11.2 μ s	11.45 μ s
Emulated VM	1.78 ms	1.8 ms
Enlightened VM	1.24 ms	1.47 ms

Figure 12: Figure 11 :

1213

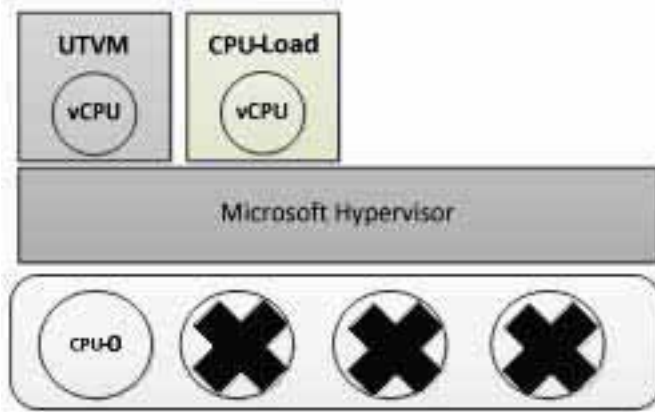


Figure 13: Figure 12 :Figure 13 :

15

Clock Tick Processing Duration	
	Maximum Overhead
Bare-Machine	10 μ s
Emulated VM	4.35 ms
Enlightened VM	1.62 ms

Figure 14: Figure 15 :

14

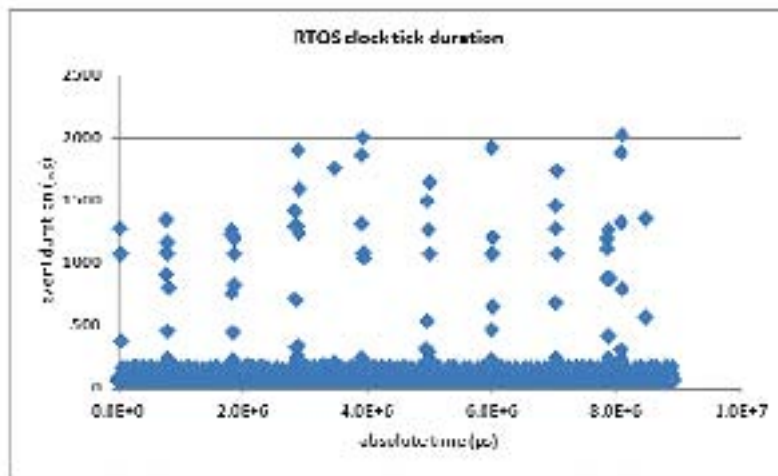


Figure 15: Figure 14 :

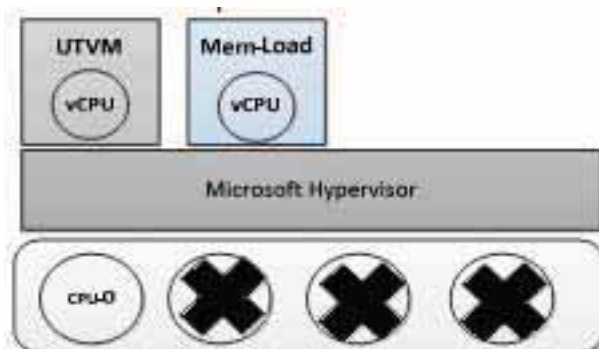


Figure 16:

2

[Note: processing duration” test results A iv. Scenario4: Contention with 1 Memory-Load VM]

Figure 17: Table 2 :

9

Figure 18: Table 9 :

10

Tables

Figure 19: Table 10 :

6

Figure 20: Table 6 :

7

[Note: test A]

Figure 21: Table 7 :

11

Figure 22: Table 11 :

12

Figure 23: Table 12 :

13

Figure 24: Table 13 :

16

Figure 25: Table 16 :

204 [Finn et al.] , P Finn , Mastering Lownds , -V Hyper , Deployment . Wiley Publishing Inc.
205 [Global Journals Inc. (US) ()] , *Global Journals Inc. (US)* 2013.

206 [Klein et al. ()] *A practitioner's Handbook for Real-Time Analysis*, M H Klein , T Ralya , B Pollak , R Obenza
207 , M G Harbour . 1994. USA: Kumer Academic Publishers.

208 [Abels et al.] *An overview of Xen Virtualization*, T Abels , P Dhawam , B Chandrasekaran . <http://www.dell.com/downloads/global/power/ps3q05-20050191-abels.pdf>
209

210 [Available] http://www.vmware.com/files/pdf/VMware_paravirtualization.pdf Available,
211 [Available] <http://msdn.microsoft.com/enus/library/cc768520%28v=bts.10%29.aspx> Available,
212 [Developers] *Buildroot: Making Embedded Linux easy*, T B Developers . <http://buildroot.uclibc.org/>
213 [CONFIG PREEMPT RT Patch-RT wiki] *CONFIG PREEMPT RT Patch-RT wiki*, (Online)

214 [Virtuatopia] *Global Journal of Computer Science and Technology Volume XIII Issue II Version I 28*, Virtuatopia
215 . <http://www.virtuatopia.com/> (An Overview of the Hyper-V Architecture)

216 [Microsoft] *Hyper-V Architecture*, Microsoft . (Online)

217 [Wiki] *Hyper-V Concepts -vCPU (Virtual Processor)*, M T Wiki . [http://social.technet.microsoft.com/wiki/contents/articles/1234.hyper-v-conceptsvcpuvirtualprocessor.aspx?wa=](http://social.technet.microsoft.com/wiki/contents/articles/1234.hyper-v-conceptsvcpuvirtualprocessor.aspx?wa=wsignin1.0)
218 [wsignin1.0](http://social.technet.microsoft.com/wiki/contents/articles/1234.hyper-v-conceptsvcpuvirtualprocessor.aspx?wa=wsignin1.0)
219

220 [Armstrong ()] *Hyper-V CPU Scheduling-Part 1*, B Armstrong . http://blogs.msdn.com/b/virtual_pc_guy/archive/2011/02/14/hyper-v-cpu-scheduling-part-1.aspx 2011.
221

222 [Corporation ()] *Hyper-V Technical Overview*, M Corporation . <http://download.microsoft.com> 2008.
223 (Windows Server)

224 [Shah ()] 'Hyper-V: Deploying Hyper-V Enterprise Server Virtualization Platform'. Z H Shah . *Windows Server*,
225 2012. 2013. Packt Publishing.

226 [Blogs] *Hyper-V: Microkernelized or Monolithic*, M T Blogs . <http://blogs.technet.com/b/chenley/archive/2011/02/23/hyper-v-microkernelized-or-monolithic.aspx>
227

228 [Microsoft ()] *Microsoft Hyper-V Server*, Microsoft . <http://www.microsoft.com/enus/server-cloud/hyper-v-server/> 2012.
229

230 [Knuth ()] *Microsoft Windows Server 2008 -Hyper-V solution overview*, G Knuth
231 . [http://www.brianmadden.com/blogs/gabeknuth/archive/2008/03/11/](http://www.brianmadden.com/blogs/gabeknuth/archive/2008/03/11/microsoft-windows-server-2008-hyper-v-solution-overview.aspx)
232 [microsoft-windows-server-2008-hyper-v-solution-overview.aspx](http://www.brianmadden.com/blogs/gabeknuth/archive/2008/03/11/microsoft-windows-server-2008-hyper-v-solution-overview.aspx) 2008.

233 [Lee et al. ()] 'Supporting Sofy Real-Time Tasks in the Xen Hypervisor'. M Lee , A S Krishnakumar , P Krishnan
234 , S Nayjot , Y Shalini . *the 6th ACM SIGPLAN/SIGOPS international conference on Virtual execution*
235 *enviroments*, 2010.

236 [Vmware ()] *Understanding Full virtualization, Paravirtualization and hardware Assist*, Vmware . 2007.