WIND RIVER

intel

# HIGH PERFORMANCE, OPEN STANDARD VIRTUALIZATION WITH NFV AND SDN

A Joint Hardware and Software Platform for Next-Generation NFV and SDN Deployments

*By John DiGiglio, Software Product Marketing, Intel Corporation*

*Davide Ricci, Product Line Manager, Wind River*

INNOVATORS START HERE.

## EXECUTIVE SUMMARY

With exploding traffic creating unprecedented demands on networks, service providers are looking for equipment that delivers greater agility and economics to address constantly changing market requirements. In response, the industry has begun to develop more interoperable solutions per the principles outlined by software defined networking (SDN) and a complementary initiative, network functions virtualization (NFV). At the heart of these two approaches is the decoupling of network functions from hardware through abstraction. The end result: Software workloads will no longer be tied to a particular hardware platform, allowing them to be controlled centrally and deployed dynamically throughout the network as needed. Moreover, network functions can be consolidated onto standard, high-volume servers, switches, and storage, further reducing time-to-market and costs for network operators.

This paper describes hardware and software ingredients addressing network equipment platform needs for NFV and SDN, and details how they could be used in a Cloud Radio Access Network (C-RAN) and other use cases. The solutions presented in this paper—designed to achieve real-time, deterministic performance using open source components—are also applicable to deploying solutions for the cloud and enterprise.

## TABLE OF CONTENTS





WIND RIVER

## KEY BENEFITS

Major network operators around the world see the potential for SDN and NFV to reduce both capital and operational expenditures (CAPEX/OPEX), as well as speed up the time-to-market for new services.

### CAPEX Savings

- **Lower hardware costs:** Take advantage of the economies of scale of the IT industry by transitioning to high-volume, industry-standard servers from purpose-built equipment that employs expensive specialty hardware components such as custom ASICs.
- **Consolidate network equipment**: Combine multiple network functions, which today require separate boxes, onto a single server (see Figure 1), thereby reducing system count, floor space, and power cable routing requirements.
- **Implement multi-tenancy:** Support multiple users on the same hardware platform, cutting down on the amount of equipment network operators need to purchase.

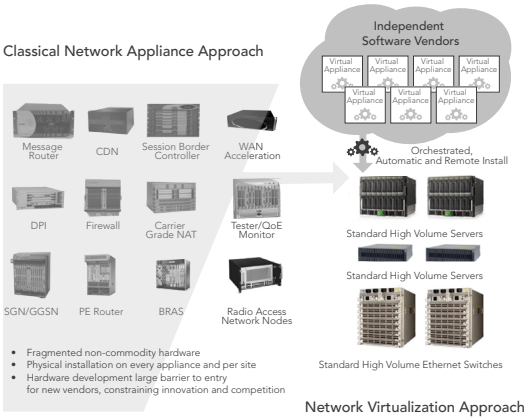


*Figure 1: From purpose-built devices to virtualized network functions running on industry-standard servers*

### OPEX Savings

- **Shorten development and test cycles:** Use virtualization to create production, test, and development sandboxes on the same infrastructure, saving time and effort.
- **Improve operational efficiency:** Simplify operations with standard servers supported by a homogeneous set of tools versus application-specific hardware with more complex, unique support requirements.
- **Reduce energy consumption:** Implement power management features available on standard servers, as well as dynamic workload rebalancing, to lower power consumption during off-peak periods.

### Service Revenue Opportunities

- **Boost innovation:** Bring new capabilities to services development while decreasing risk for network operators by enlisting an ecosystem of independent software vendors (ISVs), open source developers, and academia on the leading edge of virtual appliances.
- **Deploy services faster:** Save weeks or months when adding new services to network nodes by copying the associated software into a virtual machine (VM) instead of procuring and installing a new network appliance.
- **Target service by geography:** Increase flexibility for service rollouts to a particular geography or customer by downloading the necessary software only to applicable servers.

## DRIVING THE OPEN SOURCE SPIRIT FORWARD

Twenty years of open source evolution has reshaped entire industries. Today, open source software is everywhere: from animated movies to supercomputers used by space programs to DNA research. Often behind the scenes, open source and its thousands of contributors are transforming the world, powering businesses, connecting people, and enhancing lives.

Intel® and Wind River® are proud to be a part of this community. In fact, Intel has been there since the very beginning, long before it was a major force. Over the last two decades, both Intel and Wind River have been leading contributors to the Linux kernel, and Intel architecture is a vital foundation for many open-source-based solutions. Both companies are also taking leadership roles in the Yocto Project™, the open source collaboration that provides standardized high-quality infrastructure, tools, and methodology to help decrease the complexity and increase the portability of Linux implementations. As an active participant in the OpenStack community, Intel drove the integration of Trusted Compute Pools, used to ensure a compute node is running software with verified measurements. Intel engineers are currently developing optimizations to facilitate remote management and integration into the orchestration infrastructure. One of the optimizations provides

the orchestration layer more information about node platform capabilities stemming from PCIe-based I/O devices. Intel is also a contributing member of OpenDaylight, a community-led, industry-supported framework for accelerating SDN adoption, fostering new innovation, reducing risk, and creating a more transparent approach to networking. To learn more about Intel and the open source community, visit http://software.intel.com/en-us/oss.

### Open Components Supporting SDN and NFV

Open source software is playing a key role in networking, communications, and cloud infrastructure, enabling a move away from expensive and inflexible proprietary solutions toward those based on more open technologies with lower cost. For example, Intel and Wind River are promoting and contributing to a wide range of open source solutions, including the following:

- **Yocto Project Linux:** Open source collaboration project that provides templates, tools, and methods to help create custom Linux-based systems for embedded products regardless of the hardware architecture
- **Kernel-Based Virtual Machine (KVM):** Full virtualization solution (including a hypervisor) for Linux on Intel architecture–based hardware
- **OpenStack:** Open source cloud computing platform for public and private clouds
- **Open vSwitch:** Production quality, multilayer virtual switch licensed under the open source Apache 2.0 license
- **OpenFlow:** One of the first standard communications interfaces defined between the control and forwarding layers (i.e., node layers) of an SDN architecture
- **OpenDaylight:** Community-led, industry-supported open source framework, including code and architecture, developed to accelerate and advance a common, robust SDN platform
- **Intel Data Plane Development Kit (Intel DPDK) Accelerated Open vSwitch (Intel DPDK vSwitch):** Fork of the open source Open vSwitch multilayer virtual switch found at http://openvswitch.org/; the Intel DPDK vSwitch re-created the kernel forwarding module (data plane) by building the switching logic on top of the Intel DPDK library to significantly boost packet switching throughput; the forwarding module runs in Linux User Space with BSD license rights

## ENHANCING OPEN SOURCE FOR SDN AND NFV

Communications service providers have stringent timing constraints for their mission-critical applications and services such as voice, video, and charging. In many cases, open source software components must be enhanced in order to satisfy the associated real-time requirements. Consequently, Intel and Wind River have been working to improve the performance of network functions running in virtualized SDN and NFV environments.

A premier example is Wind River Open Virtualization Profile, an add-on to Wind River Linux 5 that provides performance enhancements, management extensions, and application services through open components. Adopting the Yocto Project as its core foundation, Wind River Linux 5 is a carrier grade, turnkey operating system that delivers all of the technologies essential to building a powerful, flexible, responsive, stable, and secure-platform. To learn more about Wind River Linux 5, please visit www.windriver.com/products/linux.

Figure 2 shows Open Virtualization Profile running along with the guest and Wind River Linux host installations. Since performance is a critical requirement, Open Virtualization Profile delivers the following:

- Real-time performance in the kernel
- Near-native application performance
- Ultra-low latency virtualization

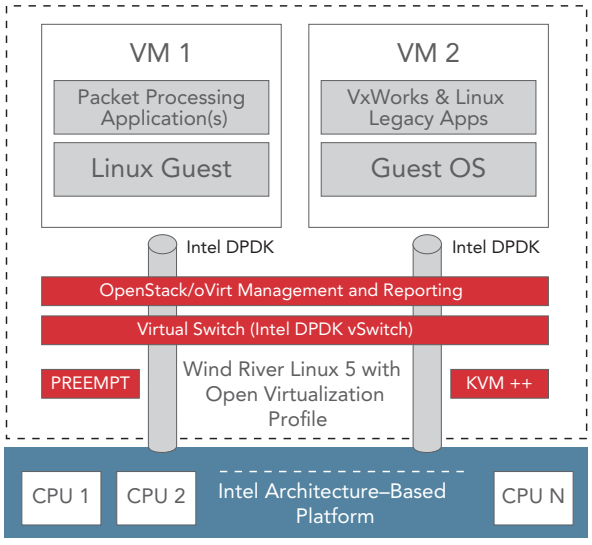


*Figure 2: Wind River Open Virtualization Profile, an add-on for Wind River Linux*

### Adaptive Performance Kernel Virtualization

Depending on the type of applications running on a system, there may be different performance requirements. For instance, throughput is of primary importance for routers; thus L3 forwarding will take precedence over most other functions. In contrast, the functions running on business support systems (BSS) platforms may all have similar priority, so the operating system may employ a round-robin or fairness approach to ensure the latency of all functions remains within an acceptable range.

Open Virtualization Profile addresses both of these circumstances with adaptive performance kernel virtualization, which makes adjustments based on the type of workload running on the system, allowing network operators to prioritize throughput or latency. The solution enables network operators to dial in performance characteristics at run time or during configuration.

- **Throughput focus:** The kernel allows for pre-emption actions that ensure real-time functions are given all the computing resources needed with minimal interruptions.
- **Latency focus:** The kernel allows the scheduler to distribute computing resources fairly; every function runs in a deterministic manner and in relative isolation from the other functions.

### Open Virtualization Profile Features

Open Virtualization Profile integrates a range of technologies and techniques to deliver adaptive performance, interrupt delivery streamlining and management, system partitioning and tuning, and security management. The adaptable, secure, and performance-oriented base software is augmented via cluster and cloud services. It supports a heterogeneous collection of hosts and guests, with options ranging from KVM guests and hosts with Wind River Linux only, through KVM guests and hosts with mixed Linux guests, to combinations of Linux and non-Linux guests. Open Virtualization Profile also produces a set of packages that can be used on non–Wind River Linux distributions, allowing integration with third-party or Wind River Linux orchestrated networks.

System-wide management and configuration technologies are provided through the integration of technologies such as libvirt, the Yocto Project's meta-virtualization layer, oVirt, and Open vSwitch. The technologies allow interoperability with public resources and the ability to interface with the resources made available on the virtual node. Application and cloud services are provided through open APIs, agents, and services that are part of an Open Virtualization Profile–powered virtual node.

## REACHING NEAR-NATIVE PERFORMANCE

It is possible to achieve near-real-time performance in SDN and NFV environments when several main issues are addressed. First and foremost, it is necessary to minimize the interrupt latency and the overhead associated with virtualized, industry-standard servers. A major source of performance loss is from VM enters and exits, which typically occur when the virtual machine monitor (VMM) must service an interrupt or handle a special event. These transitions are expensive operations because execution contexts must be saved and retrieved, and during this time the guest is stalled.

Figure 3 depicts the VM/host enters and exits following an external interrupt. In this case, the guest runs until an external interrupt arrives. Subsequently, there are a total of eight exits and enters before the guest is allowed to restart its stalled process. This overhead can become substantial since it is not uncommon for I/O-intensive applications, such as base stations, to have hundreds or thousands of interrupts arriving in a second. Similarly, a KVM guest may need to take thousands of VM exits per second because of the internal timer interrupt. These constant disruptions cannot be tolerated with communications applications because of the resulting degradation in performance, latency, and determinism.

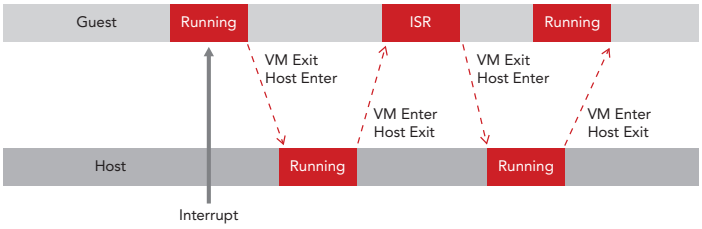


*Figure 3: Interrupt impact*

Wind River and Intel have worked together to reduce the typical interrupt latency from between 300 and 700 µS to sub-20 µS, thus achieving near-native performance (i.e., similar to non-virtualized) in a virtualized environment.

This is possible because the Wind River Open Virtualization Profile software works in conjunction with Intel Virtualization Technology (Intel VT)[1] to minimize the interrupt overhead inherent in a virtualized environment.

Some of the software mechanisms implemented in Open Virtualization Profile on top of Intel VT include:

### Guest Isolation

Open Virtualization Profile provides a high-priority guest with isolation so it can run uninterrupted and have preferential access to the hardware platform (CPU, memory, I/O devices, etc.). If the guest needs to access KVM hypervisor services or use global services, it voluntarily cedes control to the hypervisor. At this time, KVM host efficiency is critical to ensure it runs for the shortest time possible.

Another key element is the ability to deterministically direct only real-time priority interrupts to high-priority guests, thus greatly minimizing VM exits and decreasing latency. Open Virtualization Profile does this by ensuring that only the real-time interrupts are sent to the guest, and that no interrupts such as inter-processor interrupts (IPIs) for global state machines, timers, or other activities (e.g., memory reclaim) run on an isolated core. Guest isolation is also coupled with core pinning and Linux containers, providing the ability to mix workloads that have competing performance metrics and requirements, such as periodic timers. These isolated and pinned VMs (and their respective applications) run without disturbing other parts of the system, and can direct global and local resources appropriately.

### Virtual Interrupt Delivery

Open Virtualization Profile enables the VMM to inject a virtual interrupt into a guest in place of an external interrupt, which has the benefit of reducing the number of VM exits from three to one. This is because the guest is allowed to acknowledge the interrupt without triggering a VM exit. Virtual interrupt delivery greatly reduces the VM exit overhead previously described, and allows guests to run continuously for longer periods of time. This can be particularly useful for the IPI, a special type of interrupt used when one VM needs to interrupt another VM, as when two virtual switches communicate with each other.

### Core Pinning

Typically, any data that can be updated by more than one guest must be locked during access to avoid race conditions. Locks in the fast path can degrade performance 20 percent or more because they essentially eliminate the benefits of simultaneous processing while the lock is being held. To deal with this, core pinning guarantees that a particular "flow," as identified by a five-tuple (e.g., IP address, port, and protocol type) or some other predetermined criteria, is always sent to the same guest for processing. This eliminates the need for sharing connection and forwarding information among guests, because each guest only needs to know about its own connections.

### NUMA Awareness

Open Virtualization Profile uses standard Linux mechanisms to control and present the non-uniform memory access (NUMA) topology visible to guests. Among various usages, this information can help an orchestrator maximize performance by ensuring processes (e.g., QEMU) impacting a VM are not scheduled across CPUs, and the VM's memory space fits within a single NUMA node and does not cross expensive memory boundaries.

### Intel Virtualization Technology (Intel VT) for IA-32, Intel 64, and Intel Architecture (Intel VT-x)

Open Virtualization Profile takes advantage of hardware-based Intel VT to improve performance and robustness by accelerating key functions of the virtualized platform. Intel VT performs various virtualization tasks in hardware, which reduces the overhead and footprint of virtualization software and boosts its performance. Open Virtualization Profile in conjunction with Intel VT also helps avoid unintended interactions between applications by ensuring one cannot access another's memory space. Some of the underlying Intel VT technology utilized by Open Virtualization Profile to minimize virtualization overhead include the following:

- **Extended Page Tables (EPT):** Under Open Virtualization Profile control, EPT allows a guest operating system to modify its own page tables and directly handle page faults. This avoids VM exits associated with page-table virtualization, which previously was a major source of virtualization overhead. With this feature shown in the right hand box of Figure 4, a separate set of page tables, called EPT, translates guest physical addresses into host physical addresses, which are needed to access memory.

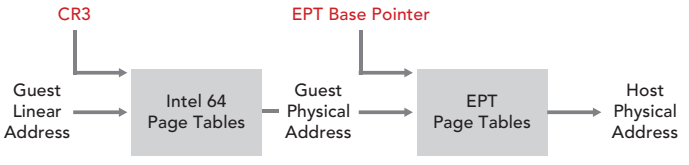


*Figure 4: EPT page tables translate guest physical addresses into host physical addresses*

- **EPT accessed and dirty bits:** The EPT has page table entries indicating whether a page was read (accessed bit) or written to (dirty bit). In addition to allowing VMs to access these bits without incurring a VM exit, these bits also enable Open Virtualization Profile to track reads/writes on memory pages in hardware, thus facilitating live migration and fault tolerance.
- **Virtual Processor IDs (VPIDs):** With VPIDs, the VM control structure contains a VM ID tag that associates cache lines with each actively running VM on the CPU. This permits the CPU to flush only the cache lines associated with a particular VM when Open Virtualization Profile performs a context switch between VMs, avoiding the need to reload cache lines for a VM that was not migrated and resulting in lower overhead.
- **Real mode support:** This feature allows guests to operate in real mode, removing the performance overhead and complexity of an emulator. Uses include:
  - Early Open Virtualization Profile load
  - Guest boot and resume

Some of the underlying Intel VT features utilized by Open Virtualization Profile to minimize interrupt latency include the following:

- **Intel VT FlexPriority:** To minimize the impact on performance, a special register called the APIC Task Priority Register (TPR) monitors in the processor the priority of tasks, to prevent the interruption of one task by another with lower priority. Intel VT FlexPriority creates a virtual copy of the TPR that can be read (see Figure 5), and in some cases changed, by guest operating systems. This eliminates most VM exits due to guests accessing task priority registers and thereby provides a major performance improvement.

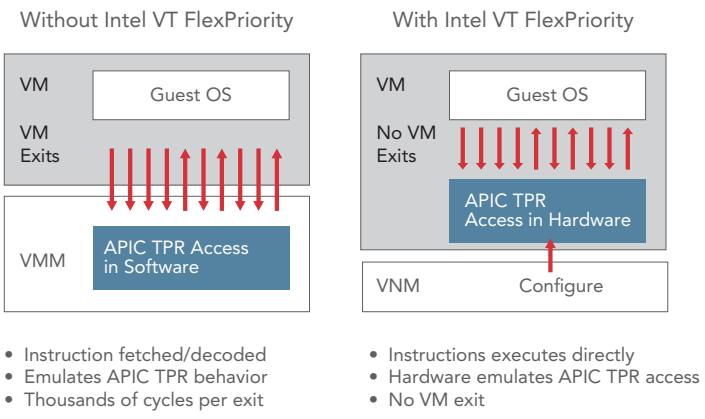


*Figure 5: Intel VT FlexPriority*

- **Guest Preemption Timer:** Programmable by Open Virtualization Profile, this timer provides a mechanism to preempt (i.e., halt) the execution of a guest operating system by causing a VM exit when the timer expires. This feature makes it easier to switch tasks, fulfill quality of service (QoS) guarantees, or allocate a certain number of CPU cycles to a task.
- **Interrupt remapping support:** This feature enables Open Virtualization Profile to isolate interrupts to CPUs assigned to a given VM and then remap or reroute the physical I/O device interrupts. When enabled, this feature helps ensure an efficient migration of interrupts across CPUs.

Intel VT also increases the robustness of virtualized environments by using hardware to prevent software running in one VM from interfering with software running in another VM, through the following technologies:

- **Descriptor table exiting:** This feature enables Open Virtualization Profile to protect a guest operating system from internal attack by preventing the relocation of key system data structures.
- **Pause-loop exiting:** Spin-locking code typically uses PAUSE instructions in a loop. This feature detects when the duration of a loop is longer than "normal" (a sign of lock-holder preemption) and forces an exit into Open Virtualization Profile. After Open Virtualization Profile takes control, it can schedule a different VM.

Open Virtualization Profile also takes advantage of Intel VT to accelerate packet movement necessary to achieve near-native application performance. These technologies include the following:

- **Address Translation Services (ATS) support:** ATS is a PCI-SIG specification that provides a mechanism for a VM to perform DMA transactions directly to and from a PCI Express (PCIe) endpoint, such as an Intel Ethernet Controller. From a high-level point of view, this is done by utilizing look-up tables to map a virtual address that the VM is accessing (reading from or writing to) to a physical location. ATS also allows a PCIe endpoint to perform DMA transactions to memory locations in a virtual machine using the same mechanism. This feature helps improve performance, since the translations can be cached at the device level, and the device need not depend on the chipset I/O translation look-aside buffer cache.

- **Intel VT for Directed I/O (Intel VT-d):** Intel VT-d accelerates data movement by enabling Open Virtualization Profile to directly and securely assign I/O devices to specific guest operating systems. Each device is given a dedicated area in system memory so data can travel directly and without Open Virtualization Profile involvement. I/O traffic flows more quickly, with more processor cycles available to run applications. Security and availability are also improved, since I/O data intended for a specific device or guest operating system cannot be accessed by any other hardware or guest software component.
- **Large Intel VT-d pages:** This feature supports 2MB and 1GB pages in Intel VT-d page tables and enables the sharing of Intel VT-d and EPT page tables.
- **Intel VT for Connectivity (Intel VT-c):** Intel VT-c performs PCI-SIG Single Root I/O Virtualization (SR-IOV) functions that allow the partitioning of a single Intel Ethernet Server Adapter port into multiple virtual functions. These virtual functions may be allocated to VMs, each with their own bandwidth allocation. They offer a high-performance, low-latency path for data packets to get into the VM. Intel VT-c enables improved networking throughput with lower CPU utilization and reduced system latency. This technology exists in Intel Ethernet NICs such as the Intel 82599 10 Gigabit Ethernet Controller.
- **Intel Data Direct I/O Technology (Intel DDIO):** Introduced with the Intel Xeon® processor E5 family, Intel DDIO allows Intel Ethernet Controllers and adapters to talk directly with the processor cache, which becomes the primary destination of I/O data (rather than main memory). The feature increases bandwidth, lowers latency, and reduces power consumption. With DDIO, the read/write operations to memory, which are very slow relative to cache memory, can be eliminated, as depicted in Figure 6. I/O-bound workloads characteristic of telecom, data plane, and network appliances can see dramatic, scalable performance benefits and reduced power consumption

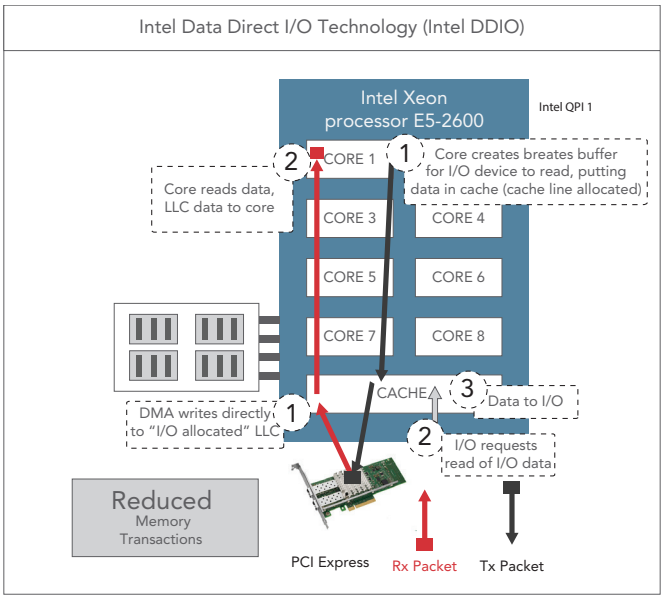


*Figure 6: Intel DDIO allows I/O devices to access the processor cache*

### Performance Results

The performance improvement delivered by Open Virtualization Profile is demonstrated by the following series of benchmark tests performed by Wind River. First, the message signaled interrupt (MSI) latency of an out-of-the box version of KVM and Linux was measured over thousands of interrupts, as shown in Figure 7a. In this virtualized environment, some interrupts had latencies exceeding 600 µs and the average was around 25 µs.

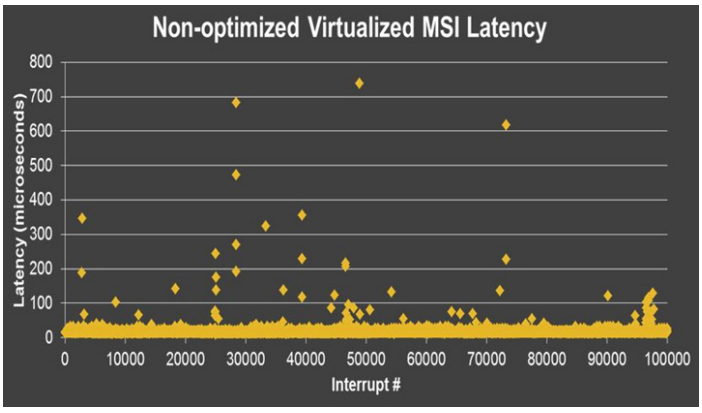


*Figure 7a: Non-optimized virtualized MSI latency*

Next, the same test was run on a system with Open Virtualization Profile, as shown in Figure 7b. The maximum interrupt latency was less than 14 µs and the average was about 8 µs. This represents a more than 40 times improvement in the worst-case latency of the non-optimized case (shown in Figure 7a), and about a three times reduction in the average interrupt latency. The optimized results using Open Virtualized Profile are similar to the non-virtualized native interrupt latency of about 10 µs for the worst case and 3 µs average, as shown in Table 1.

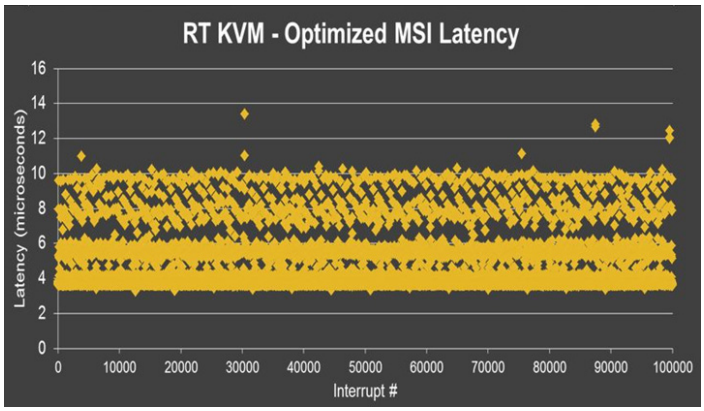


*Figure 7b: Optimized virtualized MSI latency*

*Table 1: Interrupt latency for three test conditions*

| Test Case | Interrupt Latency | |
|---|---|---|
| | Maximum (µs) | Average (µs) |
| No virtualization (native) | 9.8 | 3 |
| Optimized, virtualized | 16.9 | 3.8 |
| Non-optimized, virtualized | 760 | 25 |

### Intel QuickAssist Acceleration Technology

As the complexity of networking and security applications continues to grow, systems need more and more computational resources for workloads, including cryptography and data compression. Intel QuickAssist Technology is designed to optimize the use and deployment of algorithm accelerators in these kinds of applications.

Intel QuickAssist Technology makes it easier for developers to integrate built-in accelerators in their designs to achieve the following:

- Decrease development time by avoiding the need to create proprietary acceleration layers for each new design, device, or appliance.
- Accelerate performance for demanding applications with specific hardware acceleration modules.
- Support migration to designs using system-on-chip (SOC) and multi-core processors.
- Choose devices and solutions that fit end users' changing business requirements without being tied to a particular accelerator.

Intel QuickAssist Technology can be implemented in a few different configurations. For communications systems that require optimal use of space and thermal dissipation, Intel QuickAssist Technology is embedded into the Intel Communications Chipset 89xx Series. In this use case, the chipset functions such as Serial ATA (SATA) ports, PCIe bus extensions, USB 2.0, boot ROM, and general purpose I/O are included directly in the chipset. For commercial systems that require the acceleration performance on an enterprise class server, Intel offers Intel QuickAssist Server Accelerator Cards (QASAC), which plug into a PCIe Gen 3 slot on a standard server. Depending on the performance desired, x8 or x16 PCIe slots can be used to add Intel QuickAssist acceleration, without any degradation in performance.

In table 2, a range of Intel QuickAssist Technology solutions are shown, from a basic 1 Gbps IPsec throughput to a maximum of 80 Gbps using four QASAC cards in a standard server. All this scalability is offered using the exact same software drivers and flexible software interfaces.

*Table 2: Intel QuickAssist Technology performance ranges*

| Minimum Number Intel Xeon Processor E5-2600 Family Cores | L3 Forwarding (64 B) | | IPsec Forwarding (1 kB) | |
|---|---|---|---|---|
| Cores | Throughput | Packet Rate | Throughput | C89xx SKUs |
| DC 8C@ 2.0 GHz | 80 Gbps | ≥120 Mpps | 80 Gbps | 4xC8920 |
| UP 8C@ 2.0 GHz | 40 Gbps | 60 Mpps | 40 Gbps | 2xC8920 |
| 4C @ 2.0 GHz | 20 Gbps | 30 Mpps | 20 Gbps | 1xC8920 |
| 4C @ 1.0 GHz | 10 Gbps | 15 Mpps | 10 Gbps | 1xC8910 |
| 2C @ 1.5 GHz | 4 Gbps | 6 Mpps | 4 Gbps | 1xC8910 |
| 1C @ 1.3 GHz | 1 Gbps | 1.5 Mpps | ≥1 Gbps | 1xC8903 |

### Intel Data Plane Development Kit (Intel DPDK)

The consolidation of data and control planes on a general purpose processor has been significantly advanced by the Intel DPDK, which greatly boosts packet processing performance and throughput. Pre-integrated with Open Virtualization Profile, the Intel DPDK provides Intel architecture–optimized libraries to accelerate L3 forwarding, yielding performance that scales linearly with the number of cores, in contrast to native Linux. The solution is supported by the Wind River development environment, further simplifying use and code debugging.

The Intel DPDK contains a growing number of libraries, whose source code is available for developers to use and/or modify in a production network element. Likewise, there are various use case examples, such as L3 forwarding, load balancing, and timers, that help reduce development time. The libraries can be used to build applications based on "run-to-completion" or "pipeline" models, enabling the equipment provider's application to maintain complete control.

In addition to hardware acceleration from Intel VT and large Intel VT-d pages (specifically, 1 GB), Intel has made the Intel DPDK software available to aid in the development of I/O intensive applications running in a virtualized environment. This combination allows application developers to achieve near-native performance (i.e., similar to non-virtualized) for small and large packet processing in a virtualized environment. For instance, packet processing applications using the Intel DPDK applications can reach up to 64 B packets for 20 Gbps line rates and higher. Figure 8 shows packets per second for various packet sizes, in virtualized and non-virtualized (i.e., native) environments.

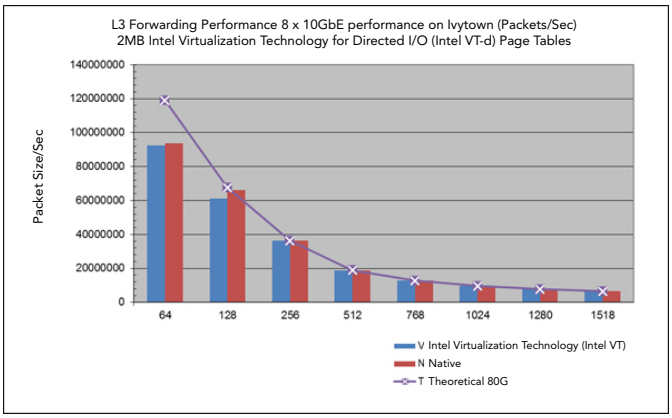


*Figure 8: Intel Data Plane Development Kit (Intel DPDK) performance*

The Intel DPDK provides a simple framework for fast packet processing in data plane applications. Developers may use the code to understand some of the techniques employed, to build upon for prototyping, or to add their own protocol stacks. SR-IOV features are also used for hardware-based I/O sharing in I/O virtualization (IOV) mode. Therefore, it is possible to partition Intel 82599 10 Gigabit Ethernet Controller NIC resources logically and expose them to a VM as a separate PCI function called a virtual function (VF). The Intel DPDK ixgbevf driver uses the NIC's virtual PCI function as a poll mode driver (PMD). Therefore, an NIC is logically distributed among multiple VMs, while still having global data in common to share with the physical function and other virtual functions.

The ixgbevf driver is added to enable inter-VM traffic using the Layer 2 switch available on the Intel 82599 10 Gigabit Ethernet Controller NIC, and consequently one can use the VF available through SR-IOV mode in the guest operating system. Inter-VM communication may take advantage of the virtual switch when VM migration is desirable, or go through the Layer 2 switch available on the NIC to optimize small packet performance.

### Open vSwitch Enhancements

Virtual switching will be a key function for many NFV deployments, and Open vSwitch is open source software capable of delivering this capability. One of the limitations of the software today is that it addresses endpoint application use where large packet sizes are typical, and is unable to switch large numbers of small packets.

Open Virtualization Profile overcomes this issue by integrating the Intel DPDK vSwitch, which takes full advantage of the the Intel DPDK high-throughput packet switching, the Intel DPDK virtualization functionality, and zero copy packet switching between switch and guest application. The Intel DPDK vSwitch also moves the software switch from the kernel to the Linux user space process, facilitating industry and proprietary enhancements.

## INTEL PLATFORM FOR COMMUNICATIONS INFRASTRUCTURE

Equipment manufacturers can economically accelerate a variety of workloads using an Intel platform that has built-in acceleration for common workloads, including packet forwarding, bulk cryptography, and compression. These capabilities, available on commercial off-the-shelf (COTS) servers, are a more flexible alternative

to purpose-built hardware. Performance throughput of 160 million packets per second (Mpps) of L3 forwarding and 80 Gbps of IPsec acceleration[2,3] have been demonstrated on servers with dual Intel Xeon processor E5-2600 series and the Intel Communications Chipset 89xx Series.

The platform includes the Intel Communications Chipset 89xx Series, which integrates SR-IOV hardware to offer Intel QuickAssist Technology accelerator services for up to 14 separate virtualized instantiations. Encryption, compression, and wireless 3G/4G LTE algorithm offload is made available to applications in individual VMs, while Intel architecture cycles are reserved for applications' general purpose compute needs.

Figure 9 illustrates the Intel Platform for Communications Infrastructure combined with the Intel 82559 10 Gigabit Ethernet Controller, Wind River Open Virtualization Profile, Intel DPDK, and Intel DPDK vSwitch to provide a high performing and robust virtualized foundation supporting SDN and NFV equipment needs.

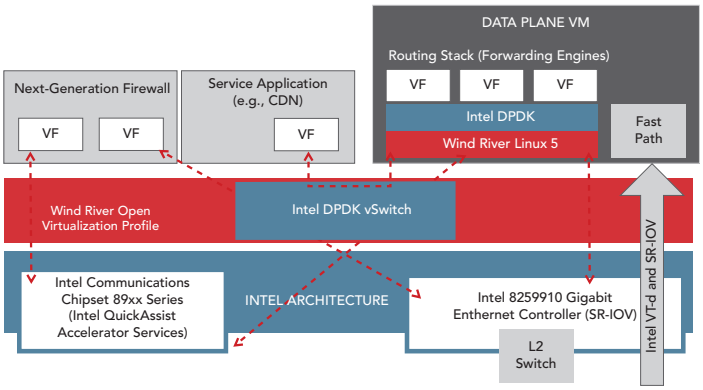


*Figure 9: Wind River Open Virtualization Profile and Intel Platform for Communications Infrastructure*

## OTHER OPEN VIRTUALIZATION PROFILE FEATURES

In addition to the mechanisms previously discussed, Wind River Open Virtualization Profile has other capabilities that are well suited for SDN and NFV deployments.

### Hot Plugging CPUs

While running real-time applications, latency must be minimized when adding, removing, or parking resources allocated to a guest. A significant issue with KVM is that it takes a relatively long time to dynamically hot plug or unplug CPUs allocated to VMs. These processes require the KVM to communicate with the guest scheduler, modify table structures, and initiate other actions that create considerable overhead.

To reallocate CPUs faster and more deterministically, Open Virtualization Profile implements dynamic resource pools that control how VMs are pinned to processor cores. For instance, presume a VM is assigned four virtual CPUs running on four physical cores; if it becomes underutilized, Open Virtualization Profile frees up two physical cores by putting all four threads on the other two physical CPUs, which can be done without performing the previously listed time-consuming tasks. Performance measurements by Wind River show it is possible to hot plug a CPU in about 40 ms and unplug a CPU in about 20 ms.

### Live Migration

Cloud infrastructure will perform live VM migration in various situations, including moving a VM to another server when its current host becomes overloaded, in order to maintain service level agreements (SLAs). Open Virtualization Profile includes migration technology that can move guests between nodes in a shelf with as little as 500 ms network downtime. This functionality can be coupled with an equipment manufacturer's other high availability mechanisms designed to perform live migration.

In addition, the capability includes various management features, such as:

- **Blacklisting:** Migration can be disabled (blocked) for applications that shouldn't migrate.
- **Reporting:** Migration failures are clearly communicated to the management interface.

### Power Management

Network operators want the ability to power down unneeded resources to save power. Open Virtualization Profile monitors resource utilization to determine when to put a node in a sleep state in order to save energy during low-use times. There are specific power governors that control power while ensuring determinism and latency specifications are met. Under the control of an orchestrator, full shutdown can be implemented as a secondary power-saving mechanism.

## VIRTUALIZATION IN THE RADIO ACCESS NETWORK

In the past five years, mobile service providers have seen an unprecedented growth in new wireless devices, subscriber applications, and cloud services. This growth is driving an unparalleled increase in traffic over service providers' networks. To support this traffic, mobile service providers need to make significant investments to their Radio Access Networks (RANs). The capital costs associated with deploying more base stations, and the operational costs associated with backhauling data from a base station to the core network, have put service providers' profitability at risk. China Mobile, one of the world's largest mobile service providers, stated that traditional RAN will become far too expensive for mobile operators to stay competitive.[4]

Figure 10 shows how 4G cellular and legacy cellular systems can be consolidated onto a single, virtualized server. In this illustration, the real-time BBU functionalities for both LTE and WCDMA run on real-time operating systems, and the non-real-time LTE and WCDMA run on other guest operating systems. This flexible platform based on Intel architecture brings the flexibility and scalability of the datacenter to the RAN.

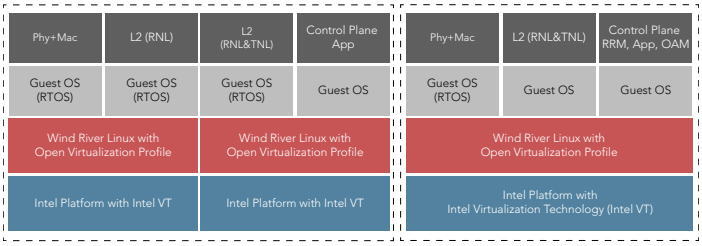


*Figure 10: Application consolidation using virtualization*

Intel and Wind River developed the C-RAN proof-of-concept using Wind River Open Virtualization Profile, as shown in Figure 11. Test results after 40 minutes and 2.5 million interrupts demonstrated that the hypervisor optimization significantly reduced latency. Table 3 shows the average latency decreased by 33 percent, and the maximum latency, which was 27 µs, by over 97 percent.[2,3,5] The optimized KVM hypervisor within Open Virtualized Profile reduced variability of the MSI interrupt latency, as well as the range between the minimum and maximum measurements. As a result, the optimized hypervisor proved to be deterministic since it satisfied 4G LTE latency and determinism requirements.

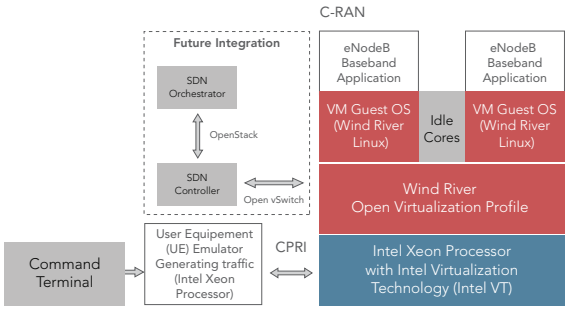


*Figure 11: C-RAN proof-of-concept*

*Table 3: MSI latency sample test results*

| | Non Optimized | Optimized | Improvement |
|---|---|---|---|
| Minimum (µs) | 10.74 | 7.65 | 28.8% |
| Maximum (µs) | 986.69 | 27.01 | 97.26% |
| Average (µs) | 18.33 | 12.18 | 33.6% |

## OTHER VIRTUALIZATION USE CASES

The following two scenarios present use cases enabled by virtualization.

### Scenario 1: Consolidating Best-of-Breed Applications with Multiple Operating Systems

Situation: An IT department wants the flexibility to choose the best VoIP and security software on the market for an appliance that is also running routing functions.

Solution: Put three workloads in separate VMs (Figure 12), allowing them to run independently on their native operating systems. As a result, IT can make application selections that are relatively independent of other software running on the system.

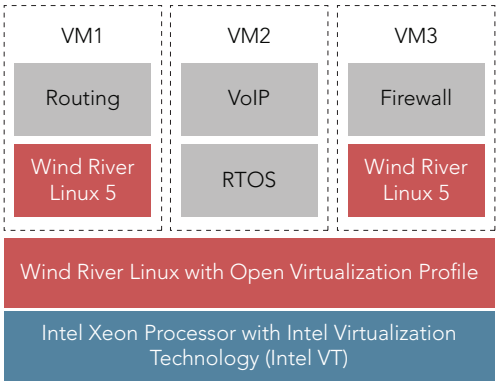


*Figure 12: Consolidating best-of-breed applications*

**Scenario 2: Application Software Isolation**

Situation: Network operators may be concerned about unintended software interactions (e.g., breaches or bugs) between applications.

Solution: Put each application into a dedicated VM, thereby isolating each execution environment and the associated data since all memory spaces are protected in hardware by Intel VT, as illustrated in Figure 13. Applications can also be assigned to dedicated processor cores in order to increase application isolation.

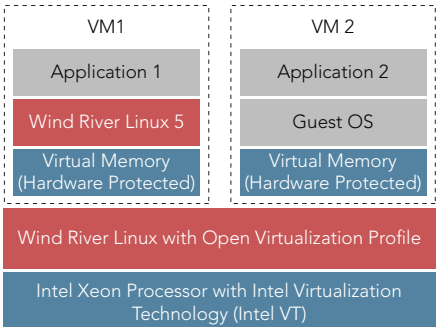


*Figure 13: Application software isolation*

## ADDING NETWORK INTELLIGENCE TO A VIRTUALIZED ENVIRONMENT

Equipment manufacturers building network elements for SDN and NFV can easily add greater network intelligence and new services to a virtualized environment by running Wind River Intelligent Network Platform. It is a comprehensive solution for the consolidation of management and data plane network applications, and contains important run-time components for data plane applications. Intelligent Network Platform integrates the Intel DPDK to offer increased packet processing performance and greater deep packet inspection (DPI) capabilities used in pattern matching, flow analysis, traffic shaping, and application identification. Operating separately or together to enable a network application running in a VM (Figure 14), there are three engines in Intelligent Network Platform:

- **Wind River Application Acceleration Engine:** This comprehensive, optimized network stack accelerates Layer 3 and 4 network protocols, networking applications, and security components.
- **Wind River Content Inspection Engine:** This high-speed engine matches large groups of regular expressions against blocks or streams in systems requiring DPI, such as intrusion prevention, antivirus and unified threat management.
- **Wind River Flow Analysis Engine:** This set of software libraries and tools enables deep visibility into layers 4–7 traffic flows, facilitating real-time packet classification, traffic categorization, and communication protocol identification, among other network applications.

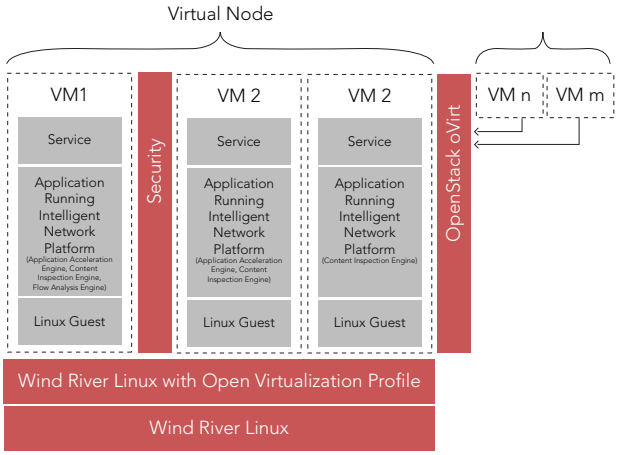


*Figure 14: Wind River Intelligent Network Platform running in a virtualized environment*

## INTEL REFERENCE DESIGN WITH OPEN VIRTUALIZATION PROFILE

Equipment manufacturers can quickly develop virtual switches for deployment in the SDN node layer by taking advantage of an Intel reference design called the Intel Open Network Platform Server Reference Design (Intel ONP Server Reference Design). As shown in Figure 15, this reference design has a node agent for OpenStack projects (Nova, Quantum, Keystone, etc.) supplied by Wind River Open Virtualization Profile, further speeding up SDN solution design. Also included is a high performance version of Open vSwitch—accelerated by the Intel DPDK—that communicates with SDN controllers. The software runs on nearly any Intel Xeon or Intel Core™ processor-based hardware platform, and combined with Intel VT, it provides a flexible, high-performance, and robust virtualization environment.

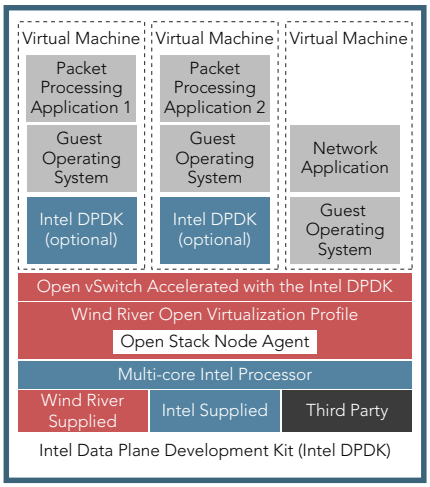


*Figure 15: Virtual switch based on the Intel ONP Server Reference Design with Open Virtualization Profile*

## CONCLUSION

To remain competitive, today's network operators must respond to evolving markets and traffic types in a timeframe of hours and days rather than the months and years more typical of traditional carrier grade networks. The network equipment platform for NFV and SDN developed by Intel and Wind River opens the door for service providers to gain unprecedented flexibility and control over customer offerings through the use of SDN and NFV.

By virtualizing network functions on an Intel and Wind River hardware and software foundation, network operators can more easily add workloads, such as DPI and power management, needed for new services and cost reduction—thereby improving the bottom line. The Wind River portfolio of embedded software solutions, including Wind River Open Virtualization Profile, combined with Intel Platform for Communications Infrastructure enables equipment manufacturers to better leverage open source components to achieve critical performance requirements, maintain maximum design flexibility, and ultimately get new products to market faster.

For more information about the Intel Platform for Communications Infrastructure, visit www.intel.com/content/www/us/en/communications/communications-overview. For more information about Wind River Open Virtualization Profile, visit www.windriver.com/announces/open_virtualization_profile or call 1-800-545-9463.

*1 Intel VT requires a computer system with an enabled Intel processor, BIOS, VMM, and, for some uses, certain platform software enabled for it. Functionality, performance, or other benefits will vary depending on hardware and software configurations and may require a BIOS update. Software applications may not be compatible with all operating systems. Please check with your application vendor.*

*2 Performance estimates are based on internal Intel analysis and are provided for informational purposes only.*

*3 Performance tests and ratings are measured using specific computer systems and/or components and reflect the approximate performance of Intel products as measured by those tests. Any difference in system hardware or software design or configuration may affect actual performance. Buyers should consult other sources of information to evaluate the performance of systems or components they are considering purchasing. For more information on performance tests and on the performance of Intel products, visit www.intel.com/performance/resources/limits.*

*4 China Mobile white paper: "C-RAN. The Road Towards Green RAN," Oct. 2011.*

*5 Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors. Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations, and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products.*

Configurations: Canoe Pass (Intel Server Board S2600CP2/S2600CP4) with 2 x Intel Xeon Processor E5-2650 @ 2.00 GHz; BIOS SE5C600.86B.01.02.0003 02/28/2012 13:35:43; 32 GB DDR3-1333 MHz; Intel Hyper-Threading Technology Disabled; Enhanced Intel SpeedStep® Technology Disabled; Processor C3/C6 Disabled; Turbo Mode Disabled; MLC Spatial Prefetcher Enabled; DCU Data Prefetcher Enabled; DCU Instruction Prefetcher Enabled; CPU Power and Performance Policy Performance; Assert NMI on SERR Disabled; Assert NMI on PERR Disabled; SMI Disabled. Software Configuration Details:(Host) Linux 3.2.14-rt24 (Host) Boot parameters: isolcpus=1-7,9-15 clocksource=tsc tsc=perfect highres=off;(Guest) Boot parameters: acpi=off root=/dev/nfs rw nfsroot=<HOST-IP>:/root/images/linux-rt-guest1-rootfs ip=dhcp isolcpus=1-3 clocksource=tsc tsc=perfect highres=off: MSI Latency testing (See test environment, this paper; unloaded, with one VM.



WIND RIVER