# Intel® Virtualization Technology for Directed I/O (VT-d): Enhancing Intel platforms for efficient virtualization of I/O devices

Submitted by **TW Burger** on Mon, 03/05/2012 - 23:16

*Virtualization solutions allow multiple operating systems and applications to run in independent partitions all on a single computer. Using virtualization capabilities, one physical computer system can function as multiple "virtual" systems.* ***Intel® Virtualization Technology*** *(Intel VT) improves the performance and robustness of today's virtual machine solutions by adding hardware support for efficient virtual machines.*

***Intel® Virtualization Technology for Directed I/O (VT-d)*** *extends Intel's Virtualization Technology (VT) roadmap by providing hardware assists for virtualization solution. VT-d continues from the existing support for IA-32 (VT-x) and Itanium® processor (VT-i) virtualization adding new support for I/O-device virtualization.*

*Intel VT-d can help end users improve security and reliability of the systems and also improve performance of I/O devices in virtualized environment. These inherently helps IT managers reduce the overall total cost of ownership by reducing potential down time and increasing productive throughput by better utilization of the data center resources.*

## Introduction

To create virtual machines (or guests) a **virtual machine monitor (VMM)** aka hypervisor acts as a host and has full control of the platform hardware. The VMM presents guest software (the operating system and application software) with an abstraction of the physical machine and is able to retain selective control of processor resources, physical memory, interrupt management, and data I/O.

A VMM supports virtualization of I/O requests from guest software. This is done in software using either of two well known models: Emulation of devices or Paravirtualization. A general reliability and protection requirement for these or any I/O-device virtualization (IOV) models is the ability to isolate and contain device accesses to only those resources that are assigned to the device by the VMM.

Intel VT-d is the latest part of the Intel Virtualization Technology hardware architecture. VT-d helps the VMM better utilize hardware by improving application compatibility and reliability, and providing additional levels of manageability, security, isolation, and I/O performance. By using the VT-d hardware assistance built into Intel's chipsets the VMM can achieve higher levels of performance, availability, reliability, security, and trust.

Intel® Virtualization Technology for Directed I/O provides VMM software with the following capabilities:

- Improve reliability and security through device isolation using hardware assisted remapping
- Improve I/O performance and availability by direct assignment of devices

## Hardware Assisted Remapping for Protection

Intel VT-d enables protection by restricting **direct memory access (DMA)** of the devices to pre-assigned domains or physical memory regions. This is achieved by a hardware capability known as DMA-remapping. The VT-d DMA-remapping hardware logic in the chipset sits between the DMA capable peripheral I/O devices and the computer's physical memory. It is programmed by the computer s ystem software. In a virtualization environment the system software is the VMM. In a native environment where there is no virtualization software, the system software is the native OS. DMA-remapping translates the address of the incoming DMA request to the correct physical memory address and perform checks for permissions to access that physical address, based on the information provided by the system software.

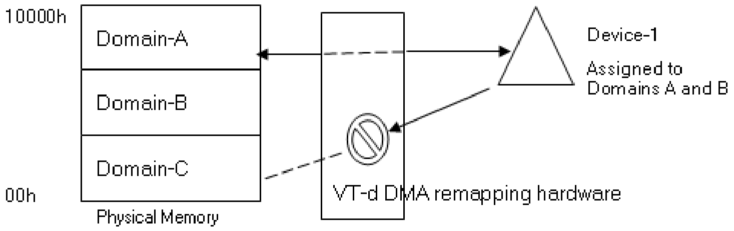
Intel VT-d enables system software to create multiple DMA protection domains. Each **protection domain** is an isolated environment containing a subset of the host physical memory. Depending on the software usage model, a DMA protection domain may represent memory allocated to a **virtual machine (VM)**, or the DMA memory allocated by a guest-OS driver running in a VM or as part of the VMM itself. The VT-d architecture enables system software to assign one or more I/O devices to a protection domain. DMA isolation is achieved by restricting access to a protection domain's physical memory from I/O devices not assigned to it by using address-translation tables. This provides the necessary isolation to assure separation between each virtual machine's computer resources.

When any given I/O device tries to gain access to a certain memory location, DMA remapping hardware looks up the address-translation tables for access permission of that device to that specific protection domain. If the device tries to access outside of the range it is permitted to access, the DMA remapping hardware blocks the access and reports a fault to the system software. Please see **Figure 1**.

**Figure-1:** VT-d DMA Remapping. Device-1 is not assigned to Domain-C, so when Device-1 tries to access Domain-C memory location range, it is restricted by the VT-d hardware.

To improve the performance, frequently used remapping-structure entries such as mapping of I/O devices to protection domains and page-table entries for DMA address translation, are cached. VT-d also supports the Peripheral Component Interconnect Special Interest Group

(PCI-SIG) Address Translation Services (ATS) specification, which specifies standard means to allow caching of device specific DMA-translations in the endpoint device.



## I/O performance through direct Assignment

Virtualization allows the creation of multiple virtual machines on a single server. This consolidation maximizes server hardware utilization, but server applications require a significant amount of I/O performance. Software based I/O virtualization methods use emulation of the I/O devices. With this emulation layer the VMM provides a consistent view of a hardware device to the VMs and the device can be shared amongst many VMs. However it could also slow down the I/O performance of high I/O performance devices. VT-d can address loss of native performance or of native capability of a virtualized I/O device by directly assigning the device to a VM.

In this model, the VMM restricts itself to a controlling function for enabling direct assignment of devices to its partitions. Rather than invoking the VMM for all (or most) I/O requests from a partition, the VMM is invoked only when guest software accesses protected resources (such as I/O configuration accesses, interrupt management, etc.) that impact system functionality and isolation.

To support direct VM assignment of I/O devices, a VMM must enforce isolation of DMA requests. I/O devices can be assigned to domains, and the DMA remapping hardware can be used to restrict DMA from an I/O device to the physical memory presently owned by its domain.

When a VM or a Guest is launched over the VMM, the address space that the Guest OS is provided as its physical address range, known as **Guest Physical Address (GPA)**, may not be the same as the real **Host Physical Address (HPA)**. DMA capable devices need HPA to transfer the data to and from physical memory locations. However, in a direct assignment model, the guest OS device driver is in control of the device and is providing GPA instead of HPA required by the DMA capable device. DMA remapping hardware can be used to do the appropriate conversion. Since the GPA is provided by the VMM it knows the conversion from the GPA to the HPA. The VMM programs the DMA remapping hardware with the GPA to HPA conversion information so the DMA remapping hardware can perform the necessary translation. Using the remapping, the data can now be transferred directly to the appropriate buffer of the guests rather than going through an intermediate software emulation layer.

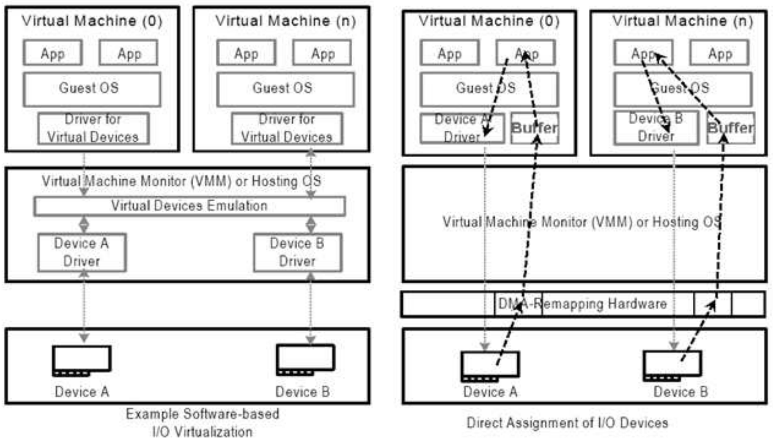


**Figure 2 - Software Emulation based I/O vs. Hardware based Direct Assignment I/O**

**Figure 2** illustrates software emulation based I/O in comparison to hardware direct assignment based I/O. In the emulation based I/O the intermediate software layer controls all the I/O between the VMs and the device. The data gets transferred through the emulation layer to the device and from the device to the to the emulation layer.

In the direct assignment model, the unmodified guest OS driver controls the device it is assigned. On the receive path the DMA-remapping hardware converts the GPA provided by the guest OS driver to the correct HPA, such that the data is transferred directly to the buffers of the guest OS (instead of passing through the emulation layer). Interrupt remapping support in VT-d architecture allows interrupt control to also be directly assigned to the VM, further reducing the VMM overheads.

## Intel VT-d Usage Models

Enabled OSs and the VMMs can utilize the VT-d functionality of I/O memory management to isolate devices to protection domains preventing devices from performing any delinquent DMA that can effect the functioning of the system.

VT-d can become the foundation for creating secure and isolated work partitions in servers, workstations and new class of combined hardware and software offerings called virtual appliances. A virtual appliance is a self-contained execution environment solution optimized to a predefined set of applications and/or services, such as a virus scanning and firewall appliance or a hardware management appliance.

Virtual machines in a virtual environment can be segregated into different protection domains from the application end to the device end. This way, a problem with one I/O device in one domain is isolated from affecting the other domains and provides IT users with better system reliability and uptime.

Test and development environments using servers with mult iple VMs, workstations with multiple co-existing OSes running in virtualized environments can all benefit from isolated work partitions.

## Server Usage Models

Many server applications are I/O intensive, especially for networking and storage. Key I/O requirements within the data center are scalability and performance. These enable server consolidation, reliability and availability as mission-critical applications are moved onto virtualized data center servers and infrastructures. Government and health care can also benefit from the isolation and security I/O virtualization provides by supporting multiple partitions with multiple OSes to meet various mission critical needs in the dynamic health care environment. Government and health care can both benefit from the added security that is the need to protect private individual's information that these institutions regularly deal with.

## Enhancing Performance

Virtualization enables the consolidation of workloads to an under-utilized server. As more work loads are consolidated the I/O usage and bandwidth requirements increase and I/O performance can become a bottleneck. To improve performance a dedicated high performance I/O device can be assigned directly to a VM that needs increased I/O performance. Intel VT-d based I/O virtualization allows high-performance I/O devices, such as multi-port gigabit and 10 gigabit network adapters, to be assigned to particular VMs where I/O performance is critical, without concerns that other VMs on the platform will affect their operation. Intel is an active participant in the PCI-SIG driven I/O virtualization specification that is working towards having a single device natively shared amongst multiple VMs.

## Enhancing Reliability and Security – Native OS and Server Consolidation

The use of multiple I/O devices in consolidated virtualized servers is increasing; up to four networking devices in a virtualized server is not uncommon. Intel VT-d can help VMMs improve reliability and security by isolating these devices to protected domains.

By controlling access of devices to specific memory ranges, end to end (VM to device) isolation can be achieved by the VMM. This helps improve security, reliability and availability.

Device isolation can be achieved in non-virtualized platforms as well. Device driver developers can use device isolation to specific memory ranges for debugging hardware or a device driver DMA that is accessing undesired memory ranges.

## Getting around "Bounce Buffer" Conditions

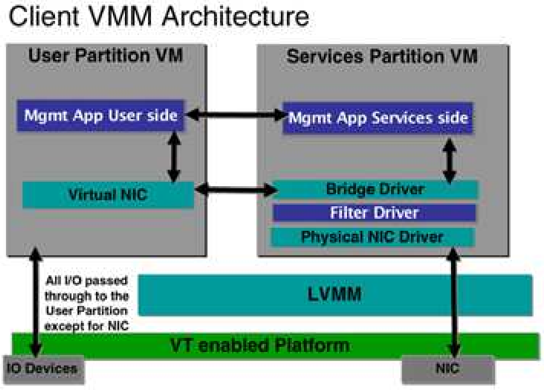
System software using Intel VT-d DMA remapping capabilities improves performance by avoiding bounce buffer**[i]** conditions. When bounce buffers are used between a 32 bit device performing DMA and a physical memory range that is inaccessible due to 32 bit address limitations, system software can use Intel VT-d DMA remapping capability to redirect the data to high memory rather than performing buffer copies.

## Client Usage Models

Intel® Virtualization Technology (Intel® VT) enables deployment of self-contained virtual appliances from third party vendors to perform vital security and management services for activities such as deep packet inspection and policy compliance on desktop PCs with Intel® vPro™ technology. These tamper resistant virtual appliances provide a more secure, stable environment for critical services and include all necessary software in a single package for greater ease and efficiency. Using VT-d with a services or manageability partition provides an isolated, controlled, and protected environment to support the client platform while assuring memory protections and I/O optimization for the virtual machines.

## VT-d based Virtual Appliances

A virtual appliance is a self-contained virtual execution environment optimized to a predefined set of applications and/or services. The **Lightweight Virtual Machine Monitor (LVMM)** is a Virtual Machine Monitor (VMM) using Intel VT to partition a client platform into two execution environments. One is the user's VM that can run an OS such as Windows XP* and applications that the user needs such as video or rendering applications, development and test applications and typical office applications. The second is a service partition (or Service VM) that runs a **services OS (SOS)** in an isolated execution environment. The user partition owns all the devices on the platform except for (in this example) the network interface controllers. These are owned by the services partition, providing an ability to monitor and/or filter network traffic and virtualize the network devices for the other VMs on the client platform. Management applications that run in the services partition provide a remote console the ability to administer the client system in isolation of the rest of the platform and user environment.



The architecture depicted in **Figure 3** shows that the network traffic flows through a physical **network interface card (NIC)** driver owned by the services partition. A bridge driver then routes the packets between the services partition network stack and the user partition network stack. In the user partition, a virtual NIC driver sends all outgoing packets from the user partition to the bridge driver and the bridge driver forwards them to the physical NIC.

**Figure 3 - Client VMM architecture**

This networking architecture provides a higher level of protection from malicious network traffic. It also creates the ability to isolate malicious attacks to a single partition and its assigned resources through the use of VT and VT-d. VT-d creates a foundation for a new class of applications based on "Virtual Appliance" architecture. It performs better than a virtualization scheme that exposes a NIC device model to the user partition. In this scheme all the user partition accesses to the NIC device are intercepted and emulated to protect proliferation of malicious code.

The LVMM and the services partition have to be protected from DMA bus mastering devices mapped to the user partition. These DMA-capable devices can access the entire system memory and can intentionally or unintentionally access (read/write) memory pages hosting the LVMM and services partition code and data structures. Such accesses could compromise IT secrets or render the platform useless by memory corruption. VT-d is used to prevent these device DMA problems.

As stated before, VT-d allows two views of the system memory: **Guest Physical Address (GPA)** and **Host Physical Address (HPA)**. The LVMM keeps the HPA view, the system physical address space and the user and services partitions are provided their respective GPA views. The LVMM maintains shadow page tables to translate GPA to HPA for accesses from the CPU. Similarly, using VT-d DMA remapping engines and corresponding translation tables, the LVMM maintains GPA-to-HPA mapping for all DMA-capable I/O devices. **Figure 4** illustrates this usage model.

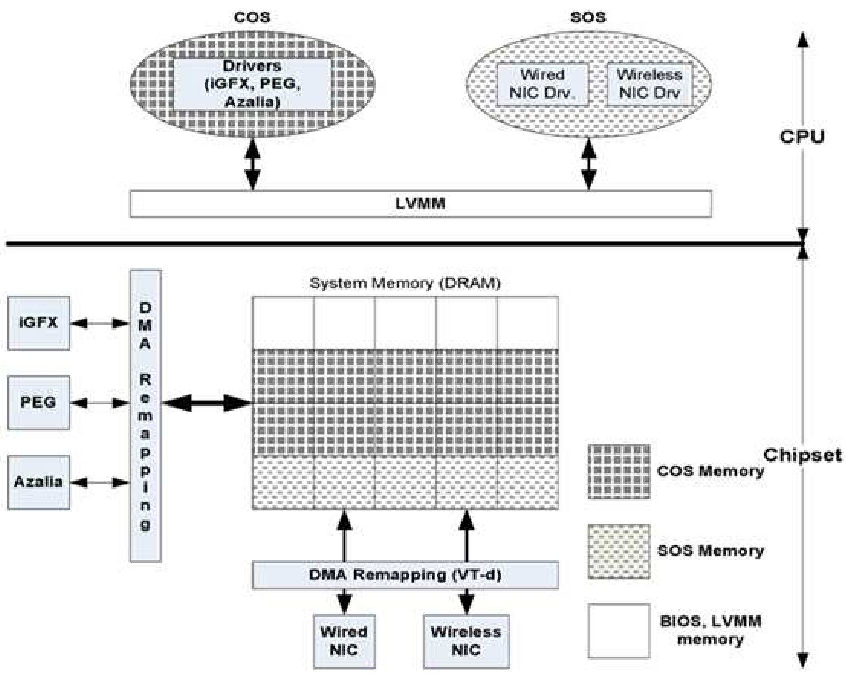


**Figure 4 - VT-d usage model in the client VMM**

DMA mapping is performed as follows:

- All services partition memory pages are added to one domain such that only DMA devices mapped to services partition (NICs) can access these pages.
- All remaining pages (except LVMM and BIOS reserved) are added to the user partition domain, and all devices except those mapped to services partition can access these pages (e.g., iGFX, PCI/PCIe add-on cards etc.).
- The LVMM and BIOS reserved regions are protected from DMA accesses by virtue of being absent from the VT-d translation page tables.

This device-to-domain mapping has the following benefits:

- I/O devices mapped to one domain can't access the memory of another domain. For example PCI/PCIe add-on cards in user partitions can't access the LVMM or the services partition.
- Device drivers in the services and user partitions run without any changes to comprehend GPA-to-HPA mapping. This translation is transparently performed by VT-d hardware when the device issues an I/O request using GPA.
- If a device misbehaves by trying to access an address outside of the mapped domain, the VT-d hardware generates a fault. This fault is captured by LVMM and is indicated to the services partition. An optional management application in the services partition can process these faults by taking appropriate actions such as displaying an error message or initiating a platform reboot, depending on the severity of the fault.

## Client Usage Models

IT departments face many issues in managing assets and, at the same time, maintaining security. Here are some examples of using VT-d in client usage models.

### Client Isolation and Recovery

IT departments benefit from the ability to isolate key manageability and security services from end-user access while still maintaining the same level of flexibility and performance for end-user services. Management and security services are isolated to a virtual management appliance or service partition, consequently, protecting the IT services. Another benefit of the user and services partitions are that if the user partition were to have a critical issue the service partition or IT partition has the ability to rebuild the user partition remotely and independently.

### Endpoint Access Control

Using VT-d to virtualize devices allows more secure **Endpoint Access Control** (**EAC** - Network Access Control). This allows more protection of client access to an enterprise creating better manageability of the access points. The enterprise determines the parameters of acceptability expressed in the form of an access policy. The policy is interpreted by a **Policy Decision Point (PDP)** which controls **Policy Enforcement Points (PEPs)** that control access. Access controls can include any of the following:

- Unrestricted access.
- Conditional access based on traffic filtering.

- Restricted access where only specific resources are accessible.

EAC follows a methodology that can be broken down into the following general steps:

- *Collection* – monitoring, reading and storage of security measurements of the client system.
- *Reporting* – formatting collected measurements for consumption by a PDP.
- *Evaluation* – interpretation of reports and organizational policies.
- *Enforcement* – applies access control rules.
- *Remediation* – applies configuration rules designed to bring the platform into compliance.

**Outbreak Containment**

IT departments continue to face the challenge of containing vulnerabilities. Viruses can enter the PC and attempt to access or harm confidential data or can proliferate throughout the enterprise. Outbreak Containment provides containment of a threat once it is detected. Intel® VT and VT-d can help detect and contain viruses sooner, limiting the exposure in attacked systems as well as other, connected systems. The Virtual Appliance or Service Partition boot process is monitored to help ensure corrupted or rogue software is not loaded if a different VMM, Virtual Appliance program, drivers, or OS attempts to load. Consequently, the corrupted partition can be halted and IT notified while allowing the uncorrupted environment, OS, and SW to load.

The corrupted partition can be switched to a private network to facilitate remediation or, in a known threat scenario; the client is updated with a patch to protect it against the outbreak. With a more serious situation, the client may be powered off to protect it and the rest of the network.

**Embedded PC Health**

Embedded PC Health reduces the client PC lifecycle costs by providing embedded asset management, provisioning, self-diagnostic, self-repair, and self-optimization capabilities within the Intel platform. This OS- independent framework, based on Intel Active Management Technology, utilizes platform-specific knowledge from Intel's processor, chipset and NIC.

Embedded PC Health main objectives are for it to be:

- *Deployable*: Utilize currently deployed protocols and services in the IT environment. Minimize the need to develop and deploy new protocols and services.
- *Highly available*: Provide remote management capabilities regardless of the operational state of the PC hardware or OS.
- *OS-independent*: Provide a base set of platform management functions and interfaces regardless of the OS type or version installed on the PC.
- *Tamper-resistant*: Prevent the end user from removing or disabling the remote management service.

## Security Implications of VT-d

Trusted partitions and memory protec tion are created on the PC to allow companies and IT to better secure sensitive data. A virtual appliance or Service Partition manages the multiple secure partitions and facilitates trusted communication of information based on business segment needs and policies. Complexity creates the potential for vulnerabilities. Using VT-d means that IT departments need to add complexity only where it is needed; creating safer execution environments and improving the ability to detect and prevent attacks.

VT and VT-d enabled systems only allow code that is approved by IT staff to be loaded. If mal-ware code is in the system, an IT verified boot procedure will detect the modification and apply the appropriate remediation such as reloading a safe backup virtual image.

Network based attacks are countered by monitoring memory pages that should not change. Monitoring agents notify the VMM when an invalid page access is attempted and the VMM can respond by blocking such accesses. The Integrity Agents are themselves protected by a VM boundary where direct access between partitions is not allowed.

IT security mechanisms are based on the ability to create isolated execution environments that are less susceptible to attack. Intel VT and VT-d technology are instrumental in creating such trusted environments that can act in the case of malicious attack or hardware failure.

## Intel® VT-d requirements

VT-d will be available on Intel Client, Workstation and select Server products in second half of 2007.

**Hardware**

- A platform that has a chipset with VT-d support

**Software enabling required for VT-d**

- A VMM (or Hypervisor) with the support required for VT-d features in the virtualization environment. No changes are required for guests running over the VMM.
- OS enabling is required for the OS to take advantage of VT-d protection features in native OS environment or non-virtualization

environment.

**BIOS requirements for the platform**

- BIOS enabling is required for VT-d use. The BIOS needs to expose VT-d capabilities (e.g. # of DMA remap engines etc) to the VMM through the ACPI table.

## Conclusion

The architecture of VT-d provides hardware mechanisms for building a virtualized environment with complete application-to-I/O device data transfer isolation. This enables the creation of a virtual environment with greater availability, reliability, and security. With VT-d, software developers can develop and evolve their architectures that provide fully protected sharing of I/O resources that are highly available, provide high performance, and scale to increasing I/O demands.

VT-d support on Intel platforms for I/O-device virtualization complements the existing Intel VT capability to virtualize processor and memory resources. Together, this roadmap of VT technologies offers a complete solution to provide full hardware support for the virtualization of Intel platforms. The virtualization of I/O resources is an important step toward enabling a significant set of emerging usage models in the data center, the enterprise, and the home.

## Resources

- Intel® Virtualization Technology - **http://www.intel.com/technology/virtualization/index.htm (http://www.intel.com/technology/virtualization/index.htm)**
- **Technology & Research (http://www.intel.com/technology/index.htm)**
- **Architecture**
- **Silicon (http://www.intel.com/technology/architecture-silicon/index.htm)**
- **Platform Benefits (http://www.intel.com/technology/product/index.htm)**
- **Software & Applications**
- **Research (http://techresearch.intel.com/articles/index.html)**
- **Standards & Initiatives (http://www.intel.com/standards/index.htm)**
- **News & Events (http://www.intel.com/design/celect/news.htm?iid=search)**
- **PCI-SIG I/O Virtualization (IOV) Specifications - Address Translation Services (http://www.pcisig.com/specifications/iov/ats)**

## Related Reading

- **How to Incorporate Intel Virtualization Technology into an Overview of Itanium Architecture**
- **How to Solve Virtualization Challenges with VT-x and VT-i**
- **Intel® Virtualization Developer Community**
- **Intel® Virtualization Technology for Directed I/O (http://www.intel.com/technology/itj/2006/v10i3/2-io/1-abstract.htm)**

## Trademark Information

Intel's trademarks may be used publicly with permission only from Intel. Fair use of Intel's trademarks in advertising and promotion of Intel products requires proper acknowledgement.

*Other names and brands may be claimed as the property of others.

**[i]** A “bounce buffer” is a memory area used for the temporary storage of data that is copied between an I/O device and a device-inaccessible memory area. This copying imposes significant overhead, resulting in increased latency, reduced throughput, and/or increased CPU load when performing I/O.

## About the Author

Thomas Wolfgang Burger is the owner of Thomas Wolfgang Burger Consulting. He has been a consultant, instructor, writer, analyst, and applications developer since 1978. He can be reached at **twburger@gmail.com**.

Categories: **Virtualization**

**For more complete information about compiler optimizations, see our Optimization Notice.**

to post comments

**RSS**

**Back to Top**

### Comments

Top

Anonymous Mon, 12/08/2008 - 02:28 to post comments

Good document about Intel VT-d.
Intel vt-d makes virtualization easier wrt performance mainly for IO intensive virtual machines. Document gives good knowledge of client usage models.

Top

Anonymous Mon, 12/08/2008 - 02:28 to post comments

Good document about Intel VT-d.
Intel vt-d makes virtualization easier wrt performance mainly for IO intensive virtual machines. Document gives good knowledge of client usage models.

Top

**Aamir Yunus (Intel)** Mon, 07/13/2009 - 10:09 to post comments

Great document!
I have a how to guide on VT-d that might be helpful for someone trying VT-d for the first time:
http://software.intel.com/en-us/blogs/2009/02/24/step-by-step-guide-on-how-to-enable-vt-d-and-perform-direct-device-assignment/

Top

Anonymous Mon, 06/28/2010 - 09:12 to post comments

Very informative document on Intel VT-d

Top

Anonymous Fri, 07/23/2010 - 18:04 to post comments

Hi Thomas,
In figure 2, the arrow from guest buffers to guest applications seems wrong. Shouldn't it be like following:
device --------> guest buffers -----> guest drivers <-------> guest application
^________________________________|

Top

Anonymous Thu, 04/28/2011 - 03:57 to post comments

missing list of chipsets that actually support the technology.
It's somewhere on Intel site but very hard to find.

Top

**TW B.** Sat, 10/05/2013 - 20:51 to post comments

Re-reading this I should have emphasized paravirtualization as the main point of VT-d. Please read:
Best Practices for Paravirtualization Enhancements from Intel® Virtualization Technology: EPT and VT-d
Submitted by Matthew Gillespie
http://software.intel.com/en-us/articles/best-practices-for-paravirtualization-enhancements-from-intel-virtualization-technology-ept-and-vt-d

Top

**TW B.** Sat, 10/05/2013 - 21:04 to post comments

Note that VT-d is a chipset Memory Controller Hub technology, not a processor feature, but this is complicated by later processor generations (i series) moving the MCH from the motherboard to the processor package, making only certain i series CPUs support VT-d.
Generally, only server chipsets are going to support VT-d since these are where virtual machines are commercially hosted. Look at http://ark.intel.com/ for servers products - chipsets

Top

**TW B.** Sat, 10/05/2013 - 21:09

to post comments

Intel® Processor Identification Utility download for Windows: http://www.intel.com/support/processors/tools/piu/sb/CS-014921.htm
Desktop Boards Compatibility with Intel® Virtualization Technology (Intel® VT) http://www.intel.com/support/motherboards/desktop/sb/CS-030922.htm

Top

Terms of Use (http://www.intel.com/content/www/us/en/legal/terms-of-use.html)

*Trademarks (http://www.intel.com/content/www/us/en/legal/trademarks.html)

Privacy (http://www.intel.com/content/www/us/en/privacy/intel-online-privacy-notice-summary.html)

Cookies (http://www.intel.com/content/www/us/en/privacy/intel-cookie-notice.html)

Publications ›

Look for us on: (http:// (http:// (http://

English ›