



SR-IOV support in Xen

Yaozu (Eddie) Dong (Eddie.Dong@intel.com)

Yunhong Jiang

Kun (Kevin) Tian



Agenda

- SR-IOV specification overview
- Xen/SR-IOV architecture
- Discussions



SR-IOV specification overview

Start with a single function device

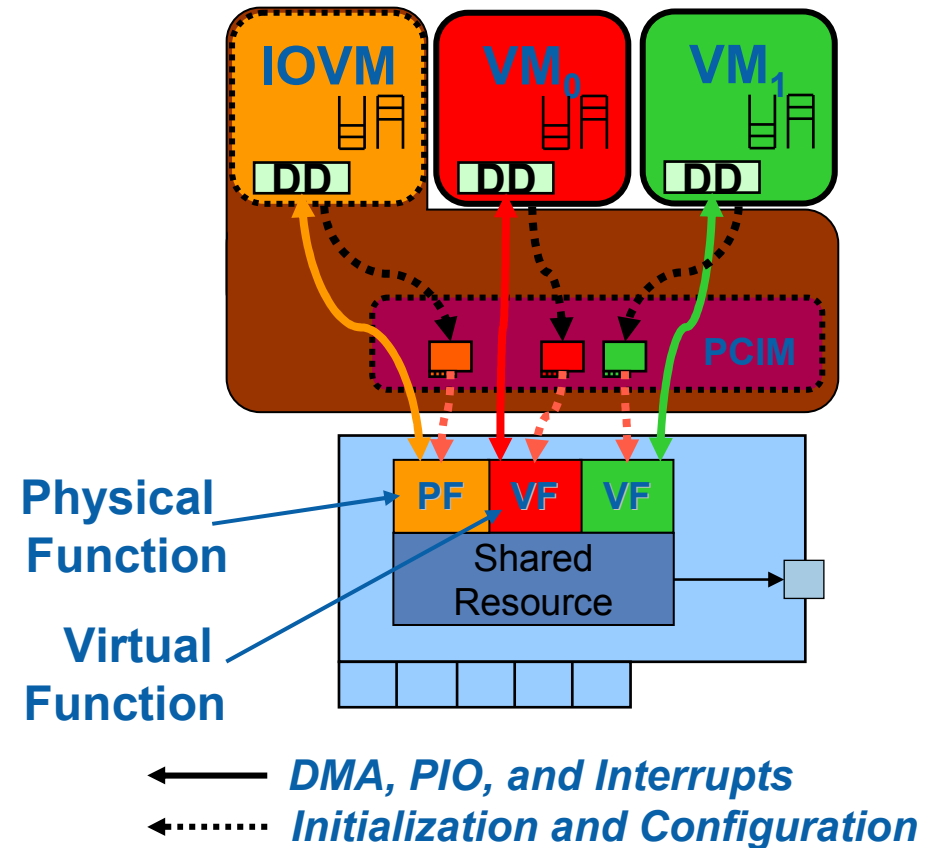
- HW under the control of privileged SW
- Includes an SR-IOV Extended Capability
- Physical Function (PF)

Replicate the resources needed by a VM

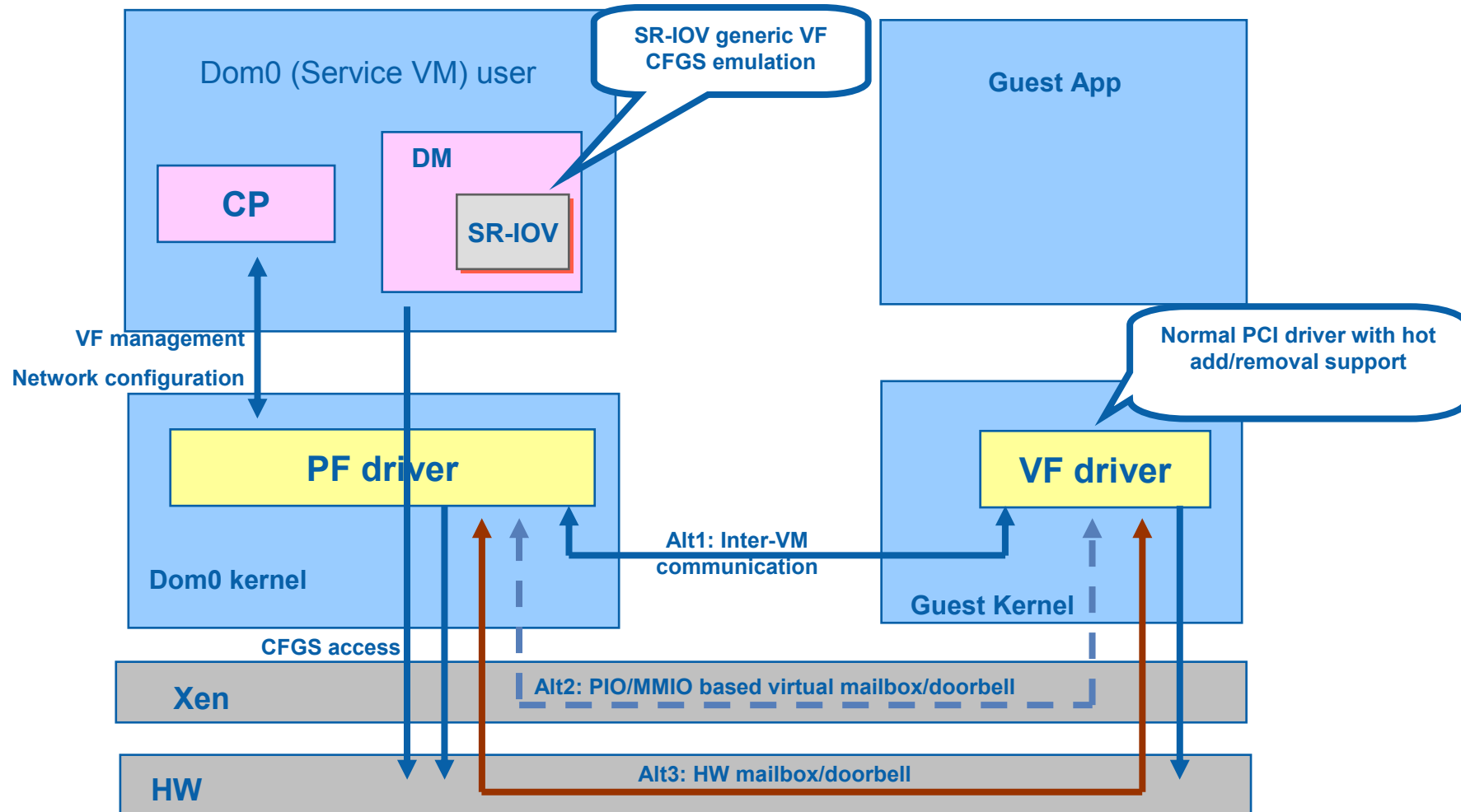
- MMIO for direct communication
- RID to tag DMA traffic
- Minimal configuration space
- Virtual Function (VF)

Introduces PCI Manager (PCIM)

- Conceptual SW entity
- Completes the configuration model
- Translates VF into a full function
- Configures SR-IOV resources



Xen/SRIOV architecture



Control Panel

VF management: A generic tool with notification to PF for state change

- **VFnums setting / Show**
- **Enable/Disable/Migrate VFs**

Network configuration

- **Can Ehtool service for this?**
- **Or a totally new tool?**



SR-IOV VF/PF Communication Channel

Inter-VM APIs → VMM cooperative VF driver

- **VF driver will depends on VMM, Guest OS, and even release from VMM vendor**
 - Primitive hypercalls are not sufficient
 - There is no Windows based Xen Inter-VM high level APIs in upstream
 - Xen vendors may release their own abstract APIs with separately

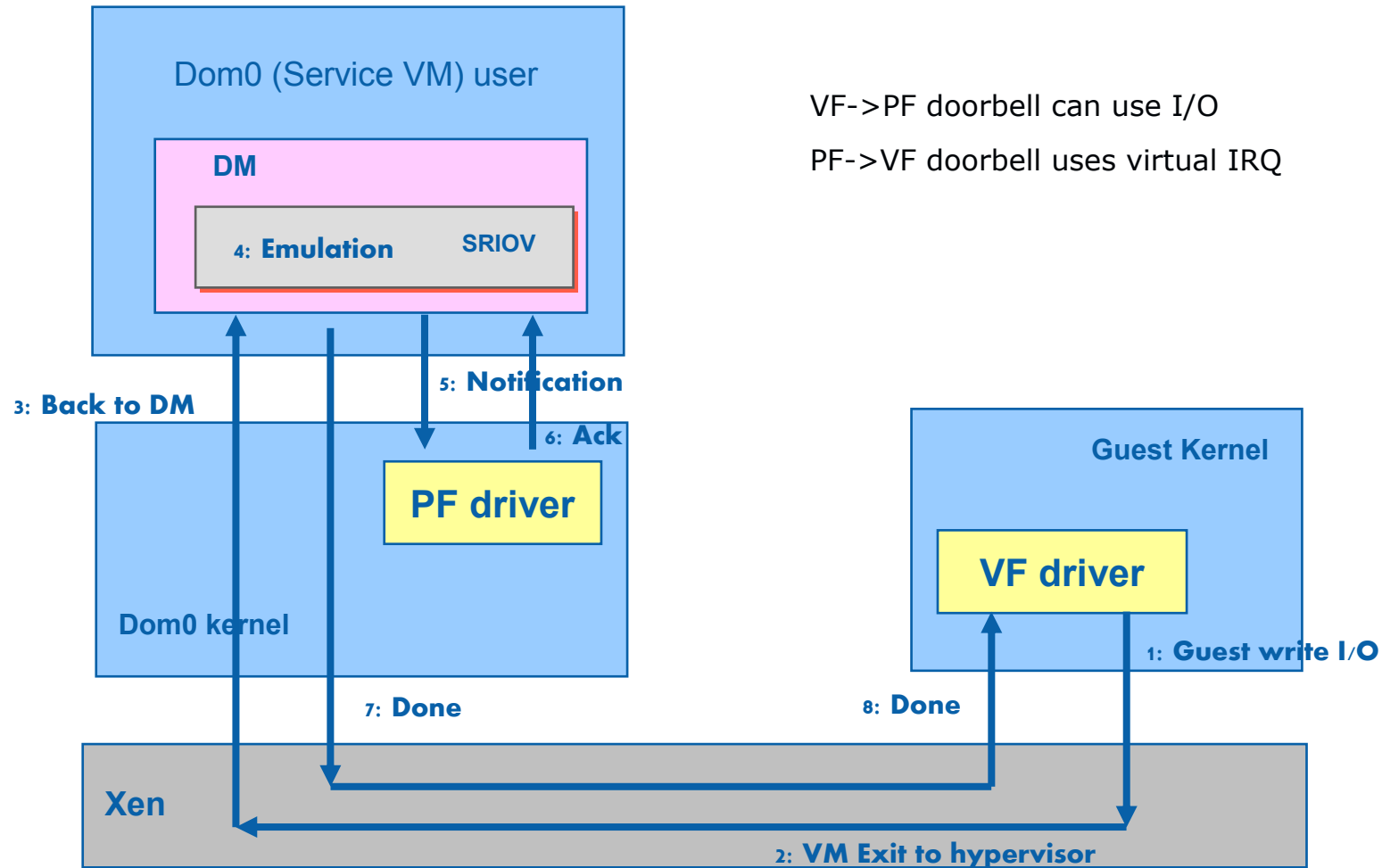
Guest hardware → VMM independent VF driver

- **Real Hardware mailbox/doorbell – No SR-PCIM involvement**
 - Good Performance, but IHVs may not implement
- **Virtual mailbox/doorbell – Need SR-PCIM support**
 - **Need SR-IOV standard**

VF/PF driver pair's decision to use whatever mechanism, but suggest using guest hardware



Virtual hardware for PF/VF communication



VF->PF doorbell can use I/O
 PF->VF doorbell uses virtual IRQ

PCI Device Instance of VF in Host?

No create VF instance

- **Need access path for Qemu to R/W VF CFGS**
- **Modifications are mostly in Qemu side**
- **Start from this for TTM**

Created VF instance

- **Pros: Easy for assignment**
- **Cons: Confuse to other pci modules, invasive change, need community decision on how to change**
- **Long term solution**



Summary

IOMMU based SR-IOV support is an ideal IO virtualization

- **Each guest can get a portion of hardware**
- **VMM doesn't need to intercept at runtime**
- **High throughput, Low CPU utilization, Perfect scalability**

Early VMM support for SR-IOV is critical to IHVs to implement PF/VF drivers

Dom0 Linux version also matters in terms of PF driver development



Software and Solutions Group



Legal Information

INFORMATION IN THIS DOCUMENT IS PROVIDED IN CONNECTION WITH INTEL® PRODUCTS. EXCEPT AS PROVIDED IN INTEL'S TERMS AND CONDITIONS OF SALE FOR SUCH PRODUCTS, INTEL ASSUMES NO LIABILITY WHATSOEVER, AND INTEL DISCLAIMS ANY EXPRESS OR IMPLIED WARRANTY RELATING TO SALE AND/OR USE OF INTEL PRODUCTS, INCLUDING LIABILITY OR WARRANTIES RELATING TO FITNESS FOR A PARTICULAR PURPOSE, MERCHANTABILITY, OR INFRINGEMENT OF ANY PATENT, COPYRIGHT, OR OTHER INTELLECTUAL PROPERTY RIGHT.

Intel may make changes to specifications, product descriptions, and plans at any time, without notice.

All dates provided are subject to change without notice.

Intel is a trademark of Intel Corporation in the U.S. and other countries.

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

Copyright © 2007, Intel Corporation. All rights are protected.



Software and Solutions Group



