Linux vSVM
Priorities and Challenges

Ashok Raj (ashok.raj@intel.com),
Yi Liu (yi.l.liu@intel.com),
Jacob Pan (jacob.jun.pan@intel.com)

Intel Corporation
September 2017


vIOMMU Architecture

Guest OS vIOMMU

Qemu/vIOMMU

VFIO API (/dev/vfio/)

User Mode Driver

VFIO Kernel Driver (core, PCI, mediated, platform dev)

VFIO type1 IOMMU

IOMMU Driver (map/unmap, IRQ remap)

VFIO PCI

VFIO mdev

VFIO platform

Device Assignment

Device

Device
Intel IOMMU SVM
vIOMMU/vSVM Priorities

✓ IOMMU provides protection between VM’s for direct assigned devices.

• vIOMMU
  – Priorities:
    • Use of DPDK inside guest.
      – gVFIO->vIOMMU->Qemu->kVFIO->IOMMU
    • vSVM – direct user mode support in guest.
    • X2APIC, Intr remapping for guests with > 255 vCPUs.
  – Others:
    • IOVA inside guest
    • IOVA for virtio devices.
Intel SVM Current Plans

• New Capability – PASID Support Limitation (PSL)
  – Only allows either a 2nd level (gIOVA), or vSVM. NOT BOTH!
  – Default: Translation in Pass through.

• Global PASID table for the system.
  – Currently they are per-iommu.

• Direct assigned devices (Covered in Yi’s/Jacob’s talk)
  – PASID table per device for SVM enabled PCIe endpoints and SR-IOV devices

• Dealing with PF/VF differences.
  – PASID & PRI – PF attributes,
  – ATS – VF attribute.
  – MAX_PASID_WIDTH. If not supporting full 20bits, treat it as if device has no pasid capability.
Opens