



THE LINUX FOUNDATION
OPEN SOURCE SUMMIT

NORTH AMERICA

VirtIO for PCI Endpoint Subsystem in Linux Kernel

Manivannan Sadhasivam

Qualcomm India Pvt Ltd



Agenda

- Overview of VirtIO
- Overview of PCI Endpoint Subsystem
- VirtIO for PCI Endpoint Subsystem
- VirtIO Proposals
- Virtio-Msg
- Future Plans



Overview of VirtIO



Overview of VirtIO

- Open standard for communication between drivers and devices



Overview of VirtIO

- Open standard for communication between drivers and devices
- Initially developed by Rusty Russell



Overview of VirtIO

- Open standard for communication between drivers and devices
- Initially developed by Rusty Russell
- Now maintained by a standards body - OASIS
 - <https://docs.oasis-open.org/virtio/virtio>



Overview of VirtIO

- Open standard for communication between drivers and devices
- Initially developed by Rusty Russell
- Now maintained by a standards body - OASIS
 - <https://docs.oasis-open.org/virtio/virtio>
- Primarily used as an I/O virtualization framework
 - Exposing I/O devices to guests by the hypervisor
 - Used by hypervisors such as KVM, lguest, ACRN etc...



Overview of VirtIO

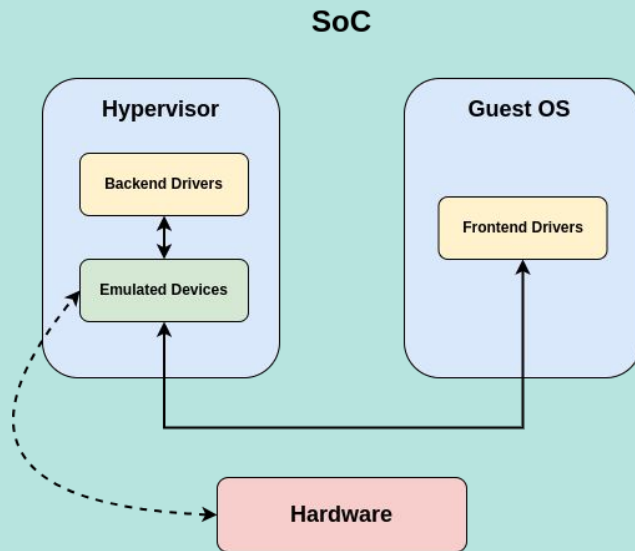
- Open standard for communication between drivers and devices
- Initially developed by Rusty Russell
- Now maintained by a standards body - OASIS
 - <https://docs.oasis-open.org/virtio/virtio>
- Primarily used as an I/O virtualization framework
 - Exposing I/O devices to guests by the hypervisor
 - Used by hypervisors such as KVM, lguest, ACRN etc...
- Also used for inter chip communication within the SoC
 - RPMSG



VirtIO Architecture



VirtIO Architecture



Overview of PCI Endpoint Subsystem



Overview of PCI Endpoint Subsystem

- Introduced in v4.12



Overview of PCI Endpoint Subsystem

- Introduced in v4.12
- Used to run Linux on PCI(e) Endpoint devices
 - NVMe, WLAN, GPU, etc...



Overview of PCI Endpoint Subsystem

- Introduced in v4.12
- Used to run Linux on PCI(e) Endpoint devices
 - NVMe, WLAN, GPU, etc...
- Components



Overview of PCI Endpoint Subsystem

- Introduced in v4.12
- Used to run Linux on PCI(e) Endpoint devices
 - NVMe, WLAN, GPU, etc...
- Components
 - PCI Endpoint Controller (PCI_EPC)
 - `drivers/pci/controllers/`



Overview of PCI Endpoint Subsystem

- Introduced in v4.12
- Used to run Linux on PCI(e) Endpoint devices
 - NVMe, WLAN, GPU, etc...
- Components
 - PCI Endpoint Controller (PCI_EPC)
 - `drivers/pci/controllers/`
 - PCI Endpoint Subsystem (PCI_EP)
 - `drivers/pci/endpoint/`



Overview of PCI Endpoint Subsystem

- Introduced in v4.12
- Used to run Linux on PCI(e) Endpoint devices
 - NVMe, WLAN, GPU, etc...
- Components
 - PCI Endpoint Controller (PCI_EPC)
 - `drivers/pci/controllers/`
 - PCI Endpoint Subsystem (PCI_EP)
 - `drivers/pci/endpoint/`
 - PCI Endpoint Function (PCI_EPF)
 - `drivers/pci/endpoint/functions/`



Overview of PCI Endpoint Subsystem

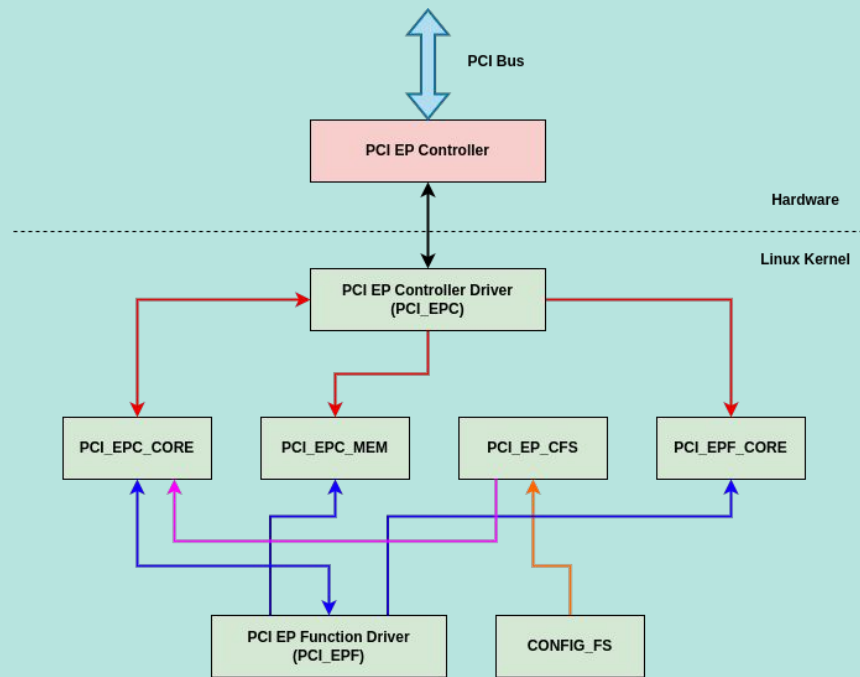
- Introduced in v4.12
- Used to run Linux on PCI(e) Endpoint devices
 - NVMe, WLAN, GPU, etc...
- Components
 - PCI Endpoint Controller (PCI_EPC)
 - `drivers/pci/controllers/`
 - PCI Endpoint Subsystem (PCI_EP)
 - `drivers/pci/endpoint/`
 - PCI Endpoint Function (PCI_EPF)
 - `drivers/pci/endpoint/functions/`
- Needs host drivers for functionality
 - `drivers/nvme/host/pci.c`



PCI Endpoint Subsystem - Architecture



PCI Endpoint Subsystem - Architecture



VirtIO For PCI Endpoint Subsystem



VirtIO For PCI Endpoint Subsystem

- Why?



VirtIO For PCI Endpoint Subsystem

- Why?
 - A Host driver is needed for each EPF driver



VirtIO For PCI Endpoint Subsystem

- Why?
 - A Host driver is needed for each EPF driver
 - Unless EPF driver follows a standard like NVMe



VirtIO For PCI Endpoint Subsystem

- Why?
 - A Host driver is needed for each EPF driver
 - Unless EPF driver follows a standard like NVMe
 - VirtIO allows reusing host drivers for many peripherals
 - virt-net, virt-scsi, virt-gpu etc...



VirtIO For PCI Endpoint Subsystem

- Why?
 - A Host driver is needed for each EPF driver
 - Unless EPF driver follows a standard like NVMe
 - VirtIO allows reusing host drivers for many peripherals
 - virt-net, virt-scsi, virt-gpu etc...
 - PCI Endpoint vendors can just develop VirtIO backend drivers



VirtIO For PCI Endpoint Subsystem

- Why?
 - A Host driver is needed for each EPF driver
 - Unless EPF driver follows a standard like NVMe
 - VirtIO allows reusing host drivers for many peripherals
 - virt-net, virt-scsi, virt-gpu etc...
 - PCI Endpoint vendors can just develop VirtIO backend drivers
 - Reduces fragmentation and lead time drastically



VirtIO Proposals



VirtIO Proposals

- SiFive
- TI
- IGEL Co., Ltd.



VirtIO Proposals

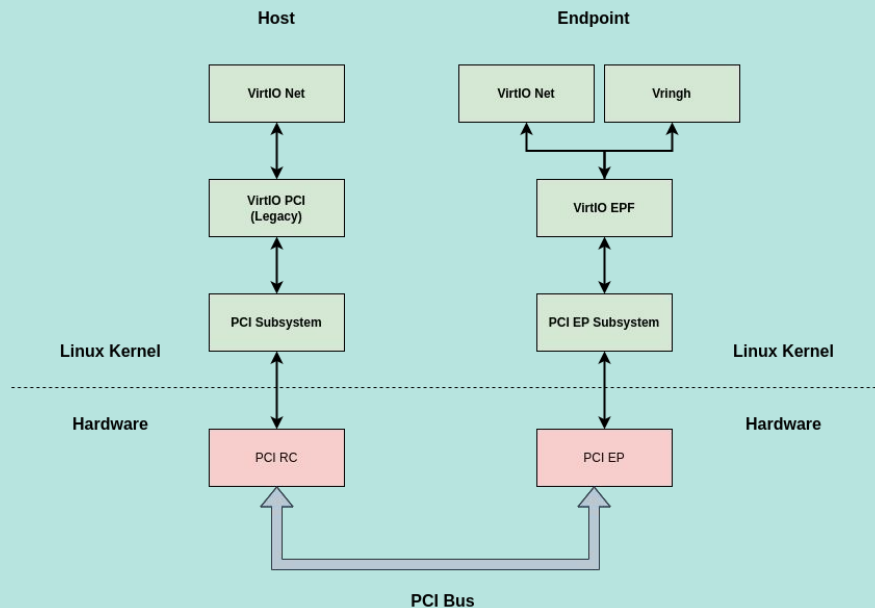
- SiFive
- TI
- IGEL Co., Ltd.
 - Selected based on consensus at LPC23



Proposed Architecture



Proposed Architecture

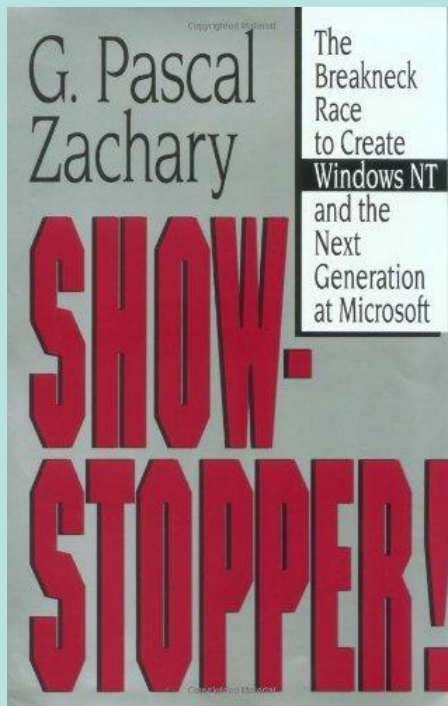


Showstopper!!!



Showstopper!!!

Not this one!



Showstopper!!!

- Implementation issues in the agreed proposal



Showstopper!!!

- Implementation issues in the agreed proposal
 - Only legacy INTx (No MSI/MSI-X)



Showstopper!!!

- Implementation issues in the agreed proposal
 - Only legacy INTx (No MSI/MSI-X)
 - Exposed legacy VirtIO device only (No modern VirtIO)



Showstopper!!!

- Implementation issues in the agreed proposal
 - Only legacy INTx (No MSI/MSI-X)
 - Exposed legacy VirtIO device only (No modern VirtIO)
 - No IOMMU support



Showstopper!!!

- Implementation issues in the agreed proposal
 - Only legacy INTx (No MSI/MSI-X)
 - Exposed legacy VirtIO device only (No modern VirtIO)
 - No IOMMU support
 - Race between VirtIO device and driver



Back to Drawing Board



VirtIO-Msg



VirtIO-Msg

- Simple message based transport for VirtIO



VirtIO-Msg

- Simple message based transport for VirtIO
 - Being developed by Linaro, ARM, Google, AMD



VirtIO-Msg

- Simple message based transport for VirtIO
 - Being developed by Linaro, ARM, Google, AMD
- <https://github.com/Linaro/virtio-msg-spec>



VirtIO-Msg

- Simple message based transport for VirtIO
 - Being developed by Linaro, ARM, Google, AMD
- <https://github.com/Linaro/virtio-msg-spec>
- Spec only defines the transport and leaves bus implementation

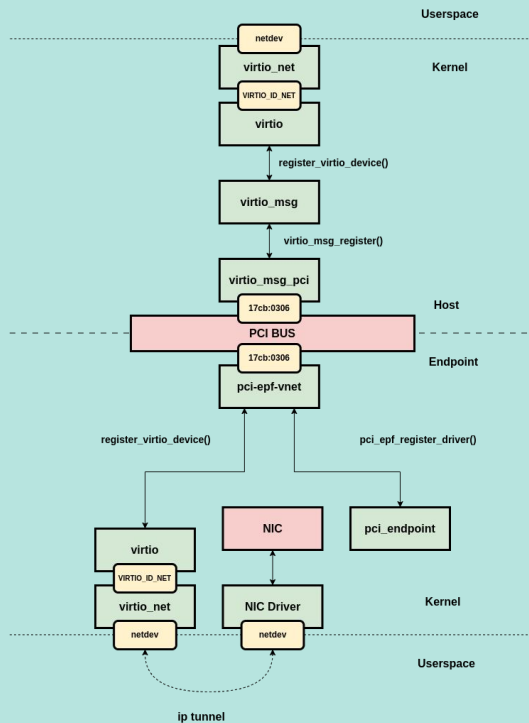


VirtIO-Msg

- Simple message based transport for VirtIO
 - Being developed by Linaro, ARM, Google, AMD
- <https://github.com/Linaro/virtio-msg-spec>
- Spec only defines the transport and leaves bus implementation
 - Ideal for PCI Endpoint Subsystem



VirtIO-Msg Architecture



VirtIO-Msg Implementation

- Bringup of virt-net took only a week!



VirtIO-Msg Implementation

- Bringup of virt-net took only a week!
- No change needed to the spec



VirtIO-Msg Implementation

- Bringup of virt-net took only a week!
- No change needed to the spec
- Able to achieve ~2Gbps iperf throughput on a PCIe Gen 3 x2 link



Future Plans



Future Plans

- Submit EPF patches once virtio-msg kernel patches are out



Future Plans

- Submit EPF patches once virtio-msg kernel patches are out
- Add more VirtIO backend drivers (virt-gpu, virt-scsi, virt-console, etc...)



Future Plans

- Submit EPF patches once virtio-msg kernel patches are out
- Add more VirtIO backend drivers (virt-gpu, virt-scsi, virt-console, etc...)
- virt-net performance improvements (RX mergeable buffer)



Future Plans

- Submit EPF patches once virtio-msg kernel patches are out
- Add more VirtIO backend drivers (virt-gpu, virt-scsi, virt-console, etc...)
- virt-net performance improvements (RX mergeable buffer)
- Encourage vendors to use VirtIO on PCI Endpoints



Thank You!

Contact me at:

mani@kernel.org

IRC: OFTC/mani_s