Contribution ID: **27** Type: **not specified**

# VFIO/IOMMU/PCI MC

The PCI interconnect specification, the devices that implement it, and the system IOMMUs that provide memory and access control to them are nowadays a de-facto standard for connecting high speed components, incorporating more and more features such as:

- Address Translation Service (ATS)/Page Request Interface (PRI)
- Single-root I/O Virtualization (SR-IOV)/Process Address Space ID (PASID)
- Shared Virtual Addressing (SVA)
- Remote Direct Memory Access (RDMA)
- Peer-to-Peer DMA (P2PDMA)
- Cache Coherent Interconnect for Accelerators (CCIX)
- Compute Express Link (CXL)
- Data Object Exchange (DOE)
- Gen-Z

These features are aimed at high-performance systems, server and desktop computing, embedded and SoC platforms, virtualization, and ubiquitous IoT devices.

The kernel code that enables these new system features focuses on coordination between the PCI devices, the IOMMUs they are connected to and the VFIO layer used to manage them (for userspace access and device passthrough) with related kernel interfaces and userspace APIs to be designed in-sync and in a clean way for all three sub-systems.

The VFIO/IOMMU/PCI micro-conference focuses on the kernel code that enables these new system features that often require coordination between the VFIO, IOMMU and PCI sub-systems.

Following up the successful LPC 2017, 2019 and 2020 VFIO/IOMMU/PCI micro-conference, the Linux Plumbers Conference 2021 VFIO/IOMMU/PCI track will therefore focus on promoting discussions on the current kernel patches aimed at VFIO/IOMMU/PCI sub-systems with specific sessions targeting discussion for the kernel patches that enable technology (e.g., device/sub-device assignment, PCI core, IOMMU virtualization, VFIO updates, etc.) requiring the three sub-systems coordination. The micro-conference will also cover VFIO/IOMMU/PCI sub-system specific tracks to debate the status of patches for the respective sub-systems.

See the following video recordings from LPC 2019 and 2020 VFIO/IOMMU/PCI micro-conference:

- VFIO/IOMMU/PCI at Linux Plumbers Conference 2019
- VFIO/IOMMU/PCI at Linux Plumbers Conference 2020

And the archived LPC 2017 VFIO/IOMMU/PCI micro-conference web page at Linux Plumbers Conference 2017, where the audio recordings from the micro-conference track and links to presentation materials are available.

The tentative schedule will provide an update on the current state of VFIO/IOMMU/PCI kernel sub-systems followed by a discussion of current issues in the proposed topics.

The following was a result of last years successful Linux Plumbers micro-conference:

- A path towards converting the Intel IOMMU driver for it to use DMA-IOMMU was defined
- Support for exposing devices to userspace using either VFIO mdev or userspace DMA was debated and brought a solution forward
- A discussion was held concerning drivers ability to enable PCI capabilities explicitly without current implicit support through the IOMMU drivers so that the number of newly added quirks can be reduced should there be a broken or buggy feature present. This discussion paved the way closer to a working solution
- The groundwork for improving security and management of both the internal and external (“trusted” and “untrusted”) devices was discussed defining changes that have to be completed going forward

- To ease problems with the hot-plug support, two concepts were presented and reviewed: movable BARs and movable bus number. A discourse followed during which the current issues were widely discussed and a possible solution was debated setting a tone for future work
- A proposal put forward to address the lack of endpoint function drivers ability to perform data transfer between the Root Complex (RC) and Endpoint (EP) leveraging the existing VirtIO infrastructure was reviewed and debated, where then a path forward has been identified
- A series of enhancements to IOMMU and VFIO user APIs for guest Shared Virtual Address (SVA) have been discussed with work already pending inclusion into the mainline kernel

Tentative topics that are under consideration for this year include (but not limited to):

- VFIO
  - Write-combine on non-x86 architectures
  - I/O Page Fault (IOPF) for passthrough devices
  - Shared Virtual Addressing (SVA) interface
  - Single-root I/O Virtualization(SRIOV)/Process Address Space ID (PASID) integration
  - PASID in SRIOV virtual functions
  - Device assignment/sub-assignment
- IOMMU
  - IOMMU virtualization
  - IOMMU drivers SVA interface
  - I/O Address Space ID Allocator (IOASID) and /dev/ioasid userspace API (uAPI) proposal
  - Possible IOMMU core changes (e.g., better integration with device-driver core, etc.)
- PCI
  - Cache Coherent Interconnect for Accelerators (CCIX)/Compute Express Link (CXL) expansion memory and accelerators management
  - I/O Address Space ID Allocator (IOASID)
  - INTX/MSI IRQ domain consolidation]
  - Gen-Z interconnect fabric
  - ARM64 architecture and hardware
  - PCI native host controllers/endpoints drivers current challenges and improvements (e.g., state of PCI quirks, etc.)
  - PCI error handling and management e.g., Advanced Error Reporting (AER), Downstream Port Containment (DPC), ACPI Platform Error Interface (APEI) and Error Disconnect Recover (EDR)
  - Power management and devices supporting Active-state Power Management (ASPM)
  - Peer-to-Peer DMA (P2PDMA)
  - Resources claiming/assignment consolidation
  - Prefetchable vs non-prefetchable BAR address mappings
  - Untrusted/external devices management
  - Thunderbolt, DMA, RDMA and USB4 security

If you are interested in participating in this micro-conference and have topics to propose, please use the Call for Proposals (CfP) process. More topics will be added based on CfP for this micro-conference.

Come and join us in the discussion in helping Linux keep up with the new features being added to the PCI interconnect specification.

We hope to see you there!

Key Attendees:

- Alex Williamson
- Benjamin Herrenschmidt
- Bjorn Helgaas
- Eric Auger
- Jason Gunthorpe
- Jean-Philippe Brucker
- Jörg Rödel
- Lorenzo Pieralisi

Contacts:


- Alex Williamson alex.williamson@redhat.com
- Bjorn Helgaas bjorn@helgaas.com
- Joerg Roedel joro@8bytes.org
- Krzysztof Wilczyński kw@linux.com
- Lorenzo Pieralisi lorenzo.pieralisi@arm.com


## I agree to abide by the anti-harassment policy

I agree


**Primary authors:** Mr WILCZYŃSKI, Krzysztof; HELGAAS, Bjorn (Google); PIERALISI, Lorenzo; WILLIAMSON, Alex; ROEDEL, Joerg (SUSE)


**Session Classification:** VFIO/IOMMU/PCI MC

**Track Classification:** VFIO/IOMMU/PCI MC