Contribution ID: **23** 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)
- Component Measurement and Authentication (CMA)
- Integrity and Data Encryption (IDE)
- Security Protocol and Data Model (SPDM)
- 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, 2020 and 2021 VFIO/IOMMU/PCI micro-conference, the Linux Plumbers Conference 2022 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, DMA ownership models, Trusted Computing, 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, 2021 and 2022 VFIO/IOMMU/PCI micro-conference:

- VFIO/IOMMU/PCI at Linux Plumbers Conference 2019
- VFIO/IOMMU/PCI at Linux Plumbers Conference 2020
- VFIO/IOMMU/PCI at Linux Plumbers Conference 2021

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:

- Support for the /dev/iommufd device has been discussed and then later refined moving it closer to the final design before implementation work would start
- The groundwork for refactoring the Shared Virtual Address (SVA) and I/O Page Fault (IOPF) support in IOMMU has been laid out readying it for future inclusion in the mainline kernel
- The SVA for in-kernel users support has taken a completely different direction following a discussion that was held concerning the proposed implementation (for instance, the KVA approach has since been retired). A new effort is to converge on the DMA API offering support through a set of possible extensions to it, however, the work is still ongoing and the final solution is yet to be decided upon
- A framework for managing group DMA ownership and removal of the BUG_ON in VFIO is now ready to be merged into the mainline kernel
- A series of enhancements for extending IOMMU domain to support more I/O Page Table types have been included in the mainline kernel
- The work to bring support for the Compute Express Link (CXL) in the Linux kernel is ongoing and it has been widely reviewed and debated, especially concerning Data Object Exchange (DOE), Component Measurement and Authentication (CMA) and Security Protocol and Data Model (SPDM) support is under heavy development. A lot of work has been put into resolving and addressing issues around the implementation, however, it will take many kernel releases before the CXL support is refined to the point where it would be considered stable to use, nonetheless good progress has been achieved thus far

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

- PCI
  - Cache Coherent Interconnect for Accelerators (CCIX)/Compute Express Link (CXL) expansion memory and accelerators management
  - Data Object Exchange (DOE)
  - Integrity and Data Encryption (IDE)
  - Component Measurement and Authentication (CMA)
  - Security Protocol and Data Model (SPDM)
  - 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
  - Probing of native PCIe controllers and general reset implementation
  - Prefetchable vs non-prefetchable BAR address mappings
  - Untrusted/external devices management
  - DMA ownership models
  - Thunderbolt, DMA, RDMA and USB4 security
- 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
    - /dev/iommufd development
    - IOMMU virtualization
    - IOMMU drivers SVA interface
    - DMA-API layer interactions and the move towards generic dma-ops for IOMMU drivers
    - Possible IOMMU core changes (e.g., better integration with device-driver core, etc.)

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
- Arnd Bergmann
- Ashok Raj
- Benjamin Herrenschmidt
- Bjorn Helgaas
- Dan Williams
- Eric Auger
- Jacob Pan
- Jason Gunthorpe
- Jean-Philippe Brucker
- Jonathan Cameron
- Jörg Rödel
- Kevin Tian
- Lorenzo Pieralisi
- Lu Baolu
- Marc Zyngier
- Pali Rohár
- Peter Zijlstra
- Thomas Gleixner

Contacts:

- Alex Williamson (alex.williamson@redhat.com)
- Bjorn Helgaas (bhelgaas@google.com)
- Jörg Roedel (jroedel@suse.de)
- Lorenzo Pieralisi (lorenzo.pieralisi@arm.com)
- Krzysztof Wilczyński (kw@linux.com)

Primary authors: WILLIAMSON, Alex (Red Hat); HELGAAS, Bjorn (Google); ROEDEL, Joerg (SUSE); WILCZYŃSKI, Krzysztof; PIERALISI, Lorenzo (Arm)

Presenters: WILLIAMSON, Alex (Red Hat); HELGAAS, Bjorn (Google); ROEDEL, Joerg (SUSE); WILCZYŃSKI, Krzysztof; PIERALISI, Lorenzo (Arm)

Track Classification: LPC Microconference Track (CLOSED)