

IOMMU Page Faulting and MM Integration

Joerg Roedel <jroedel@suse.de>

Why do we need MM Integration?

- New PCI Hardware with PRI and PASID support
 - Fault recovering
 - Multiple address spaces per device
- Allows devices to directly access process address spaces
- Translation happen in the IOMMU
- IOMMU driver needs to setup mappings for the devices

Current State

- Hardware available from AMD
 - IOMMUv2
 - Newer Radeon GPU in APUs
- Support implemented in the AMD IOMMUv2 driver
 - Extension module to the built-in AMD IOMMU driver
 - Implements the page-fault loop for devices
- Currently pending an MMU notifier extension to fix an outstanding issue
 - Will send new version when 3.18-rc1 is out

Required MMU_Notifier Change

- MMU_Notifiers not suitable for page-table sharing
 - We need the remote-TLB flush event
 - All mmu_notifiers provide is invalidate_range_start/end
 - Wrong semantics
- Patch set under review to add an invalidate_range notifier
 - Will close this gap
 - Notifies about the remote-TLB flush event
 - Also notifies when page-table pages are freed

Future

- More Hardware with these capabilities is coming up
 - Intel SVM already specified in the Vt-d specification
 - Other architectures, non-pci?
- The existing AMD code needs to be turned into a generic IOMMU-API extension
- Code is mostly generic already, the call-backs into the in-kernel AMD IOMMU driver needs to be generalalized
- Revisit PASID allocaiton/handling (David?)

Current Exported API

- `int amd_iommu_init_device(struct pci_dev *pdev, int pasids)`
- `void amd_iommu_free_device(struct pci_dev *pdev)`
- `int amd_iommu_bind_pasid(struct pci_dev *pdev, int pasid,
 struct task_struct *task)`
- `void amd_iommu_unbind_pasid(struct pci_dev *pdev, int pasid)`
- `int amd_iommu_set_invalid_ppr_cb(struct pci_dev *pdev,
 amd_iommu_invalid_ppr_cb cb)`
- `int amd_iommu_set_invalidate_ctx_cb(struct pci_dev *pdev,
 amd_iommu_invalidate_ctx cb)`



Corporate Headquarters
Maxfeldstrasse 5
90409 Nuremberg
Germany

+49 911 740 53 0 (Worldwide)
www.suse.com

Join us on:
www.opensuse.org

Unpublished Work of SUSE LLC. All Rights Reserved.

This work is an unpublished work and contains confidential, proprietary and trade secret information of SUSE LLC. Access to this work is restricted to SUSE employees who have a need to know to perform tasks within the scope of their assignments. No part of this work may be practiced, performed, copied, distributed, revised, modified, translated, abridged, condensed, expanded, collected, or adapted without the prior written consent of SUSE. Any use or exploitation of this work without authorization could subject the perpetrator to criminal and civil liability.

General Disclaimer

This document is not to be construed as a promise by any participating company to develop, deliver, or market a product. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. SUSE makes no representations or warranties with respect to the contents of this document, and specifically disclaims any express or implied warranties of merchantability or fitness for any particular purpose. The development, release, and timing of features or functionality described for SUSE products remains at the sole discretion of SUSE. Further, SUSE reserves the right to revise this document and to make changes to its content, at any time, without obligation to notify any person or entity of such revisions or changes. All SUSE marks referenced in this presentation are trademarks or registered trademarks of Novell, Inc. in the United States and other countries. All third-party trademarks are the property of their respective owners.

