Netdev 0x17

IDPF Live Migration Support

Yahui Cao, Phani R Burra, Anjali Singhai, Sridhar Samudrala

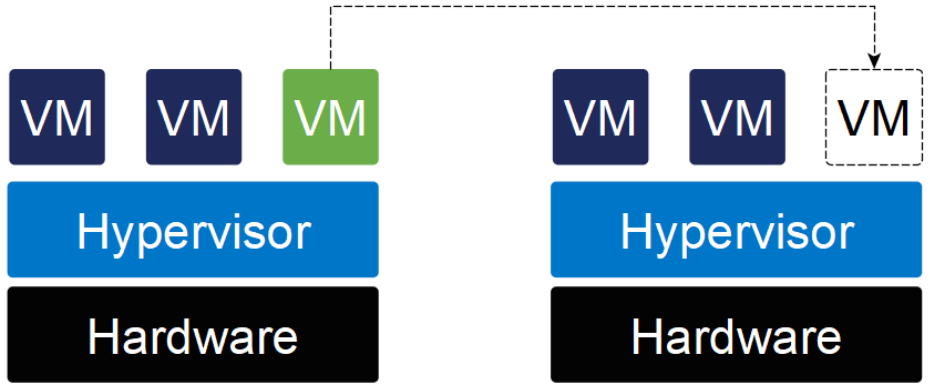
Agenda

- Overview
- Feature Discovery
- Device State
- DMA Logging
- Other

Overview

Live Migration definition:

- **moving a running VM** between different physical machines **w/o disconnecting the client**

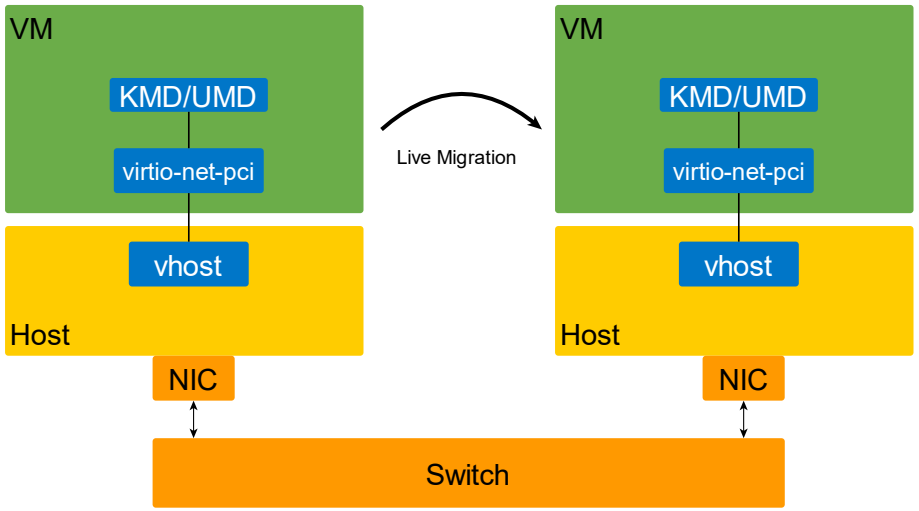


Use cases:

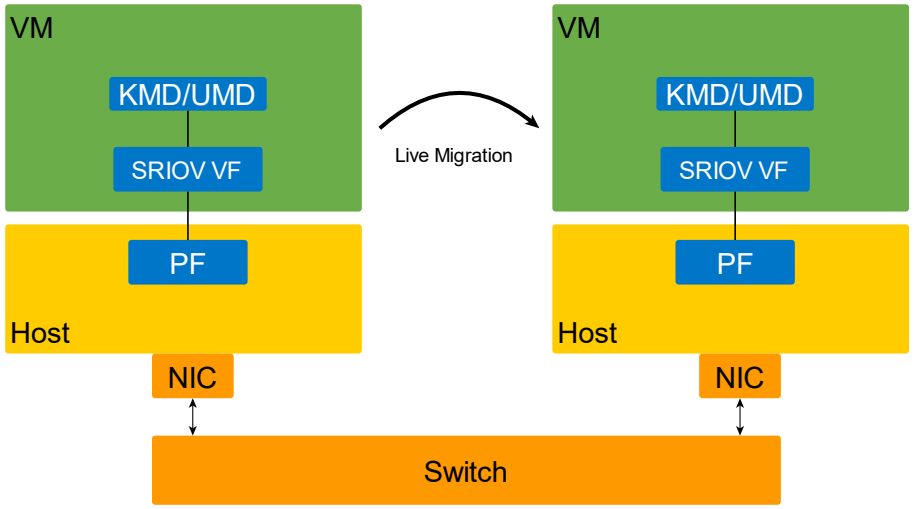
- Host software upgrades
- Load Balancing
- HW maintenance

Overview

- Status quo: Virtio vs VFIO



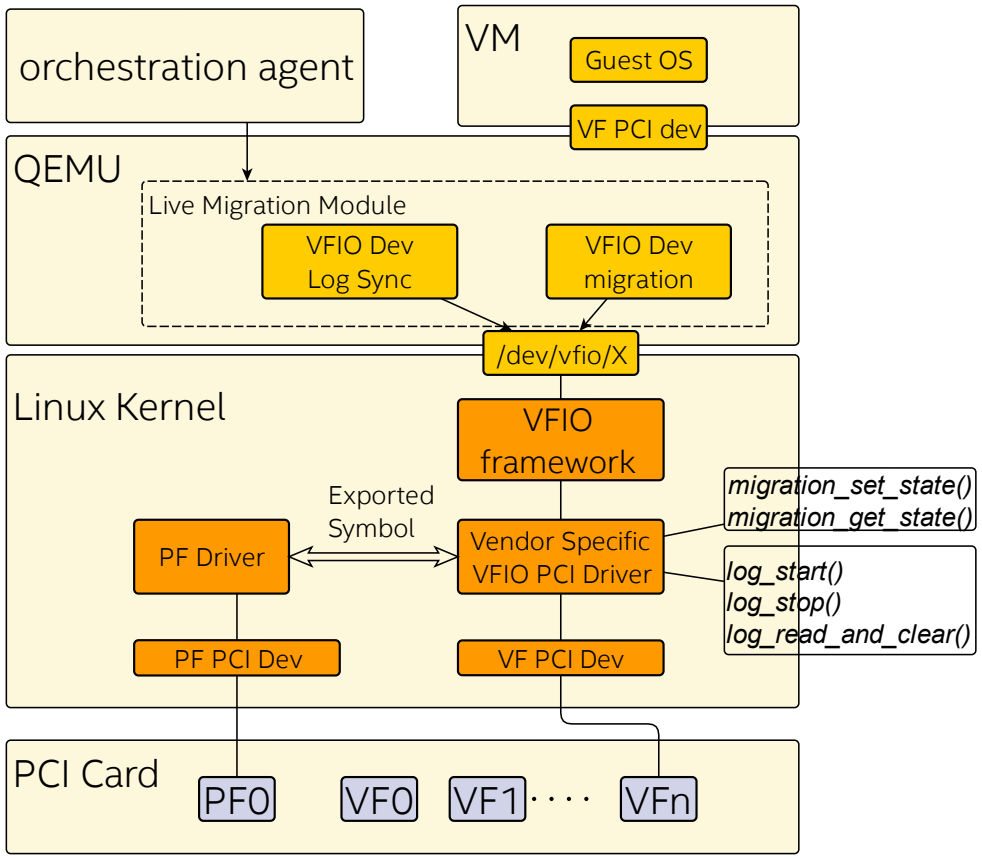
Virtio migration



VFIO migration

Overview

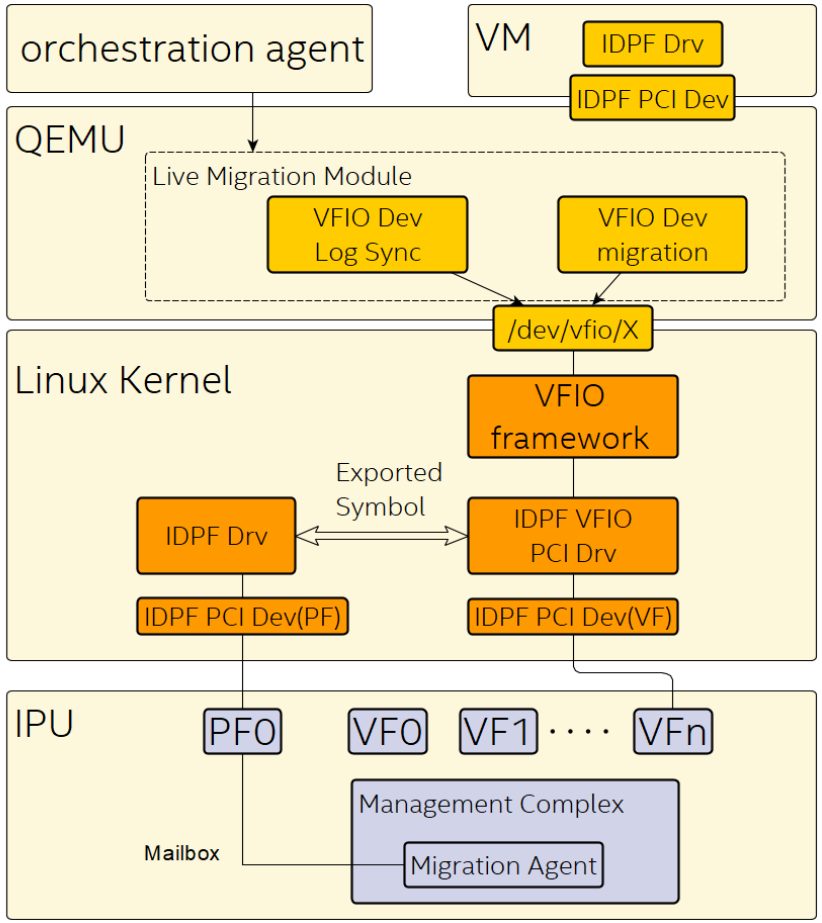
- Community framework for VFIO migration
- VFIO migration uAPI
  - Extend VFIO device uAPI
  - Save/Load/Suspend/Resume
  - Log Start/Stop/Query
- Vendor specific VFIO PCI driver
  - Register migration callbacks
  - Register DMA logging callbacks
- PF driver exports migration helper function



Overview

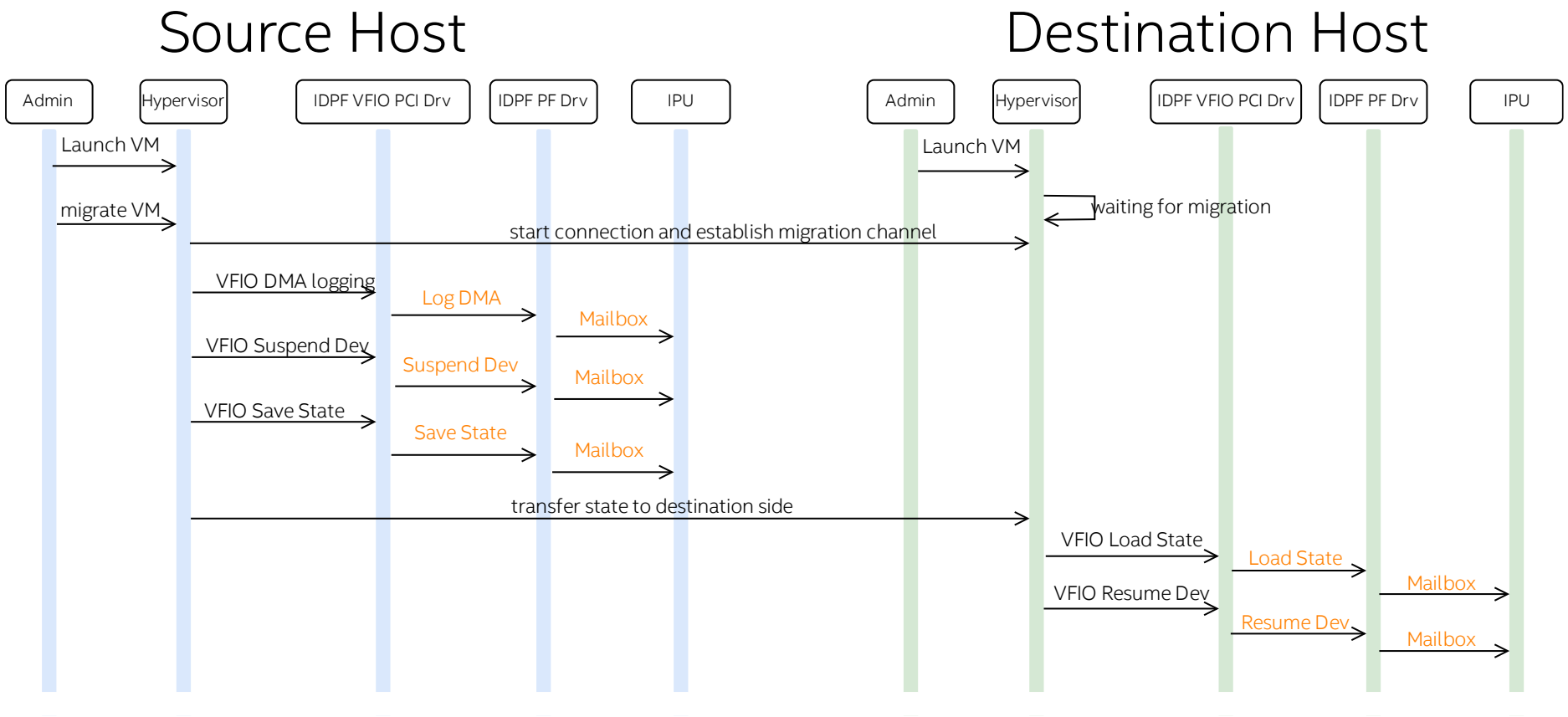
IDPF Live Migration

- IDPF PF driver exports LM helper
- IDPF VFIO PCI driver registers callbacks
- Mailbox/Virtchnl2.0 extended for migration
- Migration agent serves IDPF request



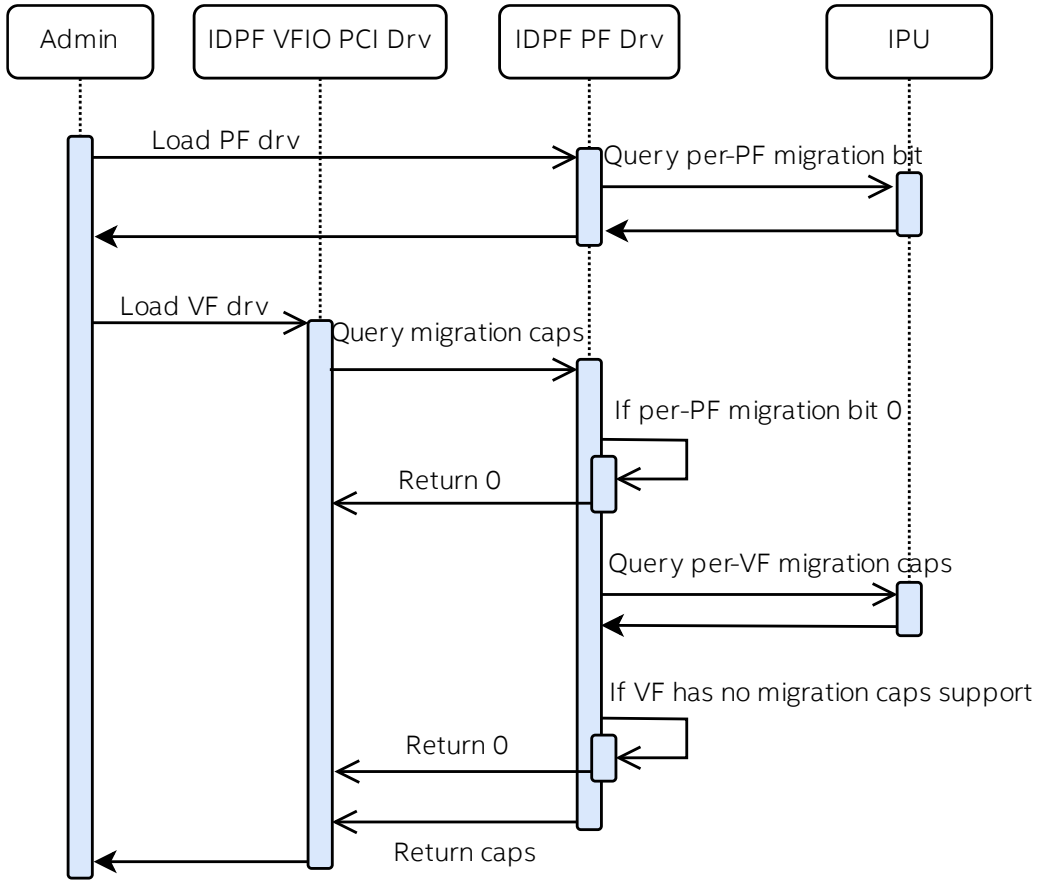
Overview

- IDPF Migration Flow



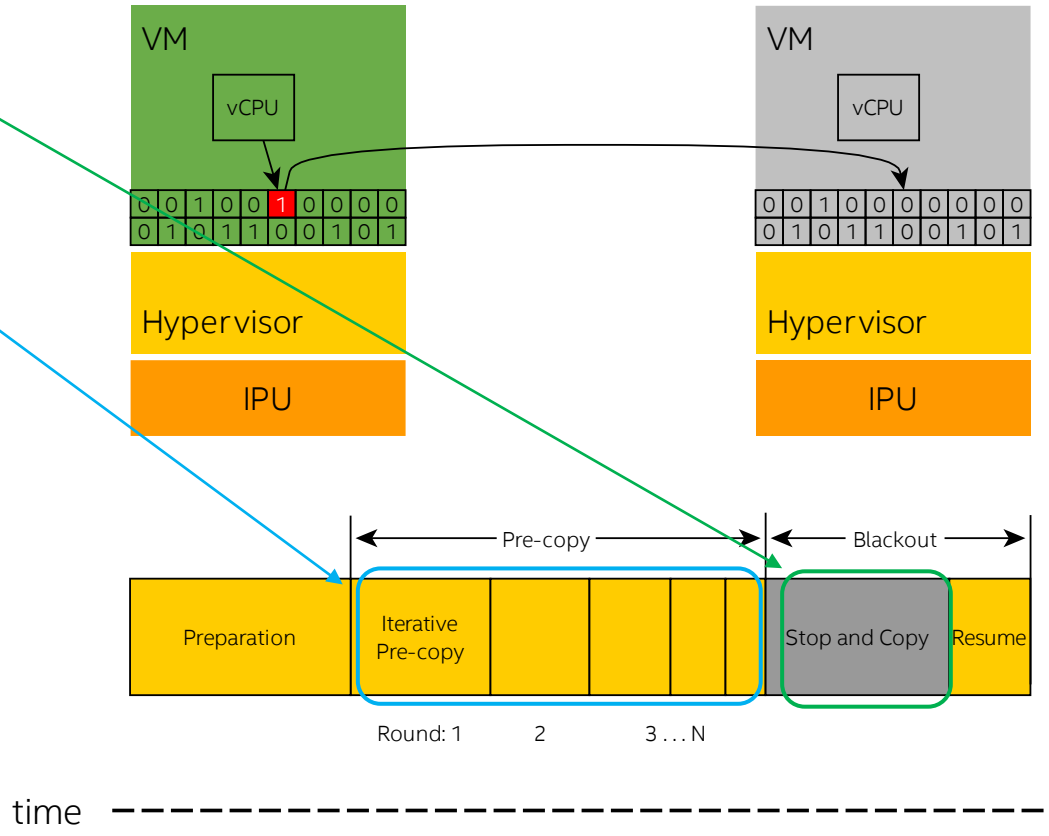
Feature Discovery

- Migration Feature Availability
  - IDPF Host PF NIC w/ SR-IOV ✓
  - IDPF Host PF NIC w/o SR-IOV ✗
  - IDPF Guest VF NIC ✗
- per-PF caps:
  - new MIGRATION bit when IDPF *Get Capabilities*
- per-VF caps: new query cmd
  - STOP_COPY: Stop dev and copy state
  - PRE_COPY: Pre-copy dev state before stop
  - DMA_LOG: Log device DMA write



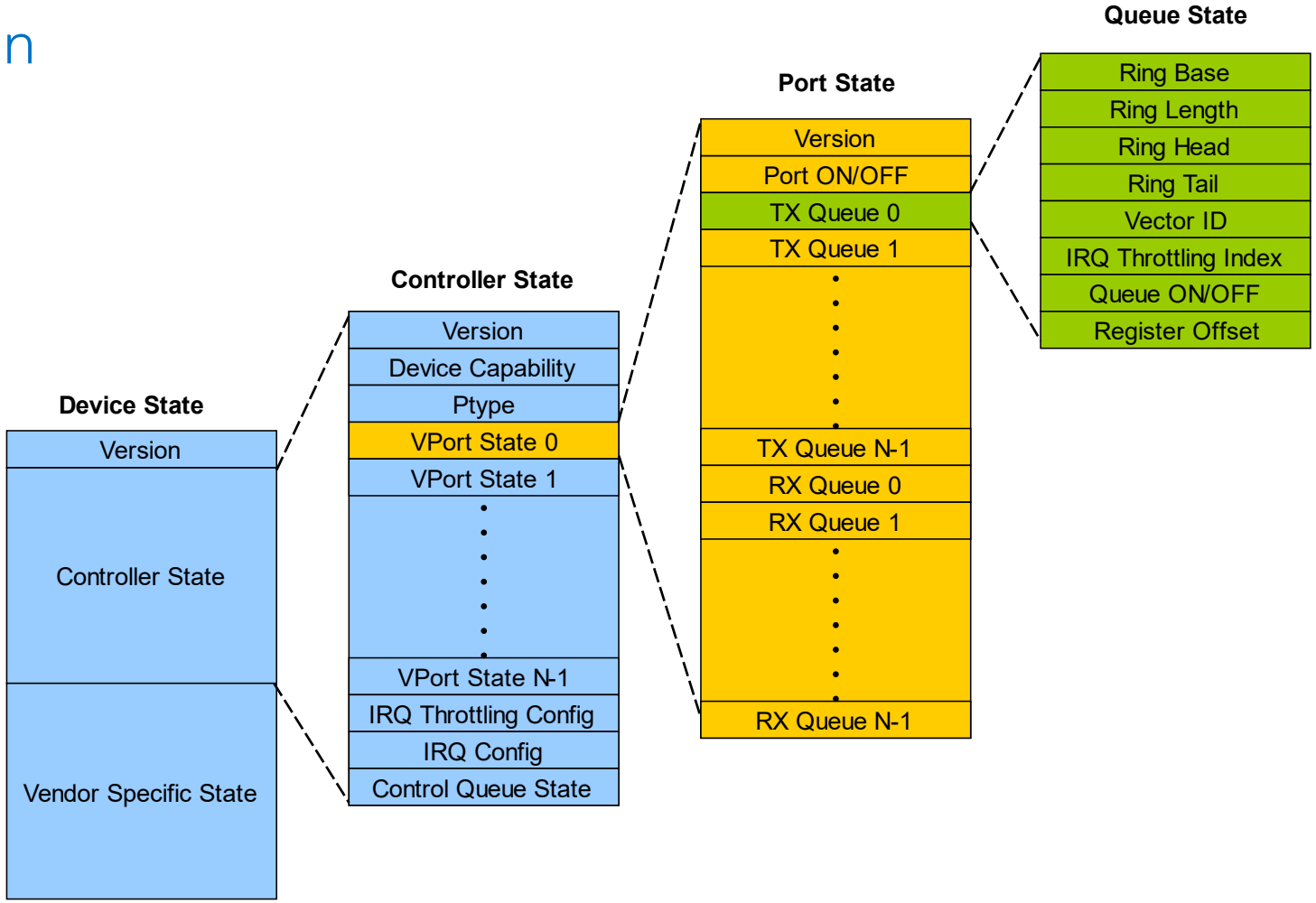
Device State

- Migration Caps: PRE_COPY, STOP_COPY
- Migration Primitive:
  - Suspend Dev, Save State, Load State, Resume Dev
- IDPF virtchnl2.0 cmd
  - *VIRTCHNL2_OP_SUSPEND/RESUME_DEV*
  - *VIRTCHNL2_OP_SAVE/LOAD_DEVSTATE*
  - *VIRTCHNL2_OP_QUERY_DEVSTATE_SIZE*



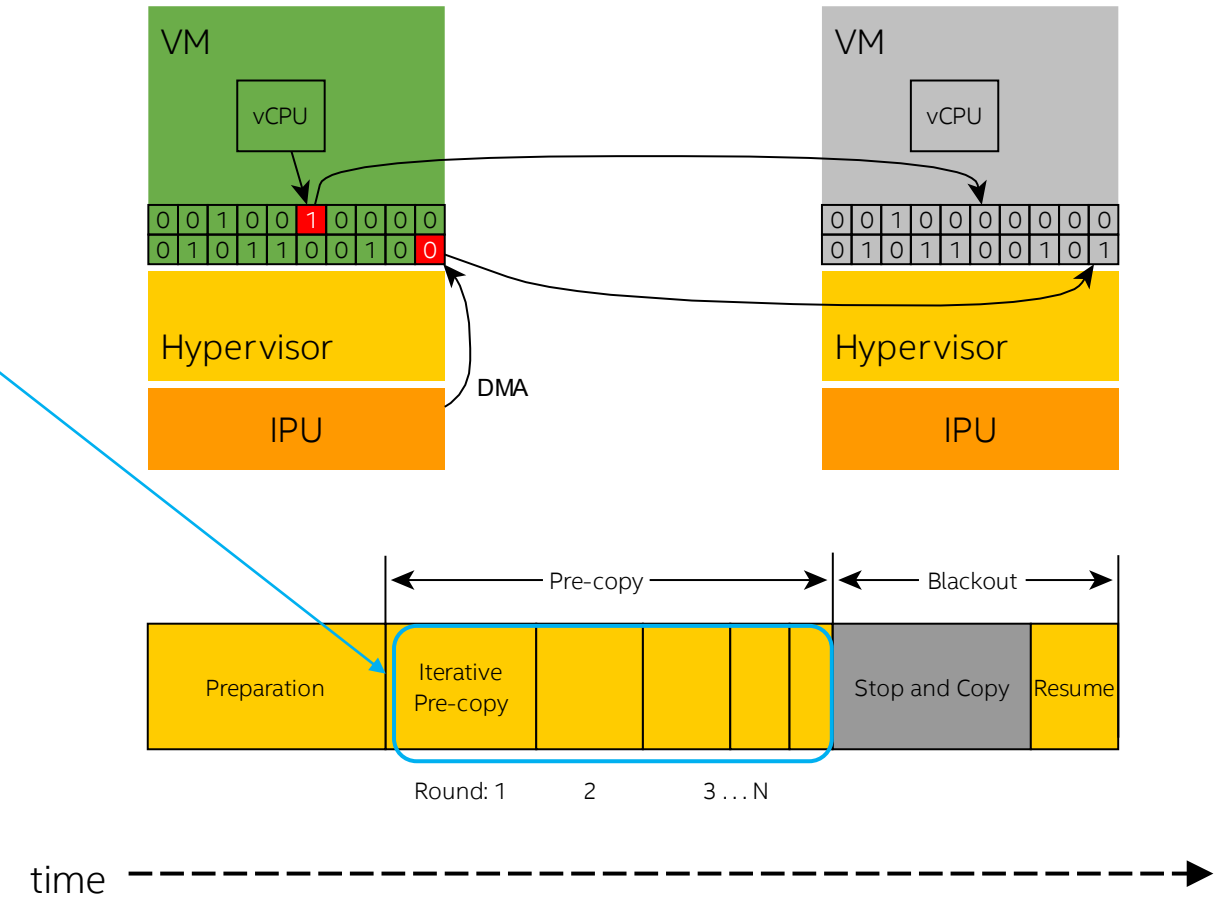
Device State

- Device state definition
  - Standardized
  - Versioned
  - Hierarchy
  - Flexible



DMA Logging

- Migration caps: DMA_LOG
- Background
  - Guest memory delta for pre-copy stage to reduce VM blackout time
  - Device DMA write bypasses hypervisor and leads to guest memory delta
- Primitive
  - Log Start/Stop/Query
- IDPF virtchnl2.0 cmd
  - *VIRTCHNL2_OP_START_DMA_LOG*
  - *VIRTCHNL2_OP_STOP_DMA_LOG*
  - *VIRTCHNL2_OP_QUERY_DMA_LOG*



Other

- vSwitch state support
- Includes: L2/L3/Tunnel rules, etc…
- Management options:
  1. Managed by Live Migration in Vendor Specific State
  2. Managed by control plane software (e.g. SDN)