EVE OS
Networking & Connectivity

Milan Lenčo, Software Engineer, ZEDEDA

THE LINUX FOUNDATION

Agenda

- **What we have today**
  - Device connectivity
  - Network instances
  - Application network adapters
- **Work in progress**
  - VLAN and LAG support
  - Hardening & Refactoring
- **R&D planned for 2022** (NFV use-cases)
  - SR-IOV support
  - DPDK integration
  - SmartNICs and HW offloading

EVE Networking
What we have today

THE LINUX FOUNDATION

Device Connectivity

- Connectivity between edge node and the controller
- Ethernet (fiber, Sat modem, etc.), WiFi, LTE
- **Unreliable**: redundancy, failover, fallback
- **Metered**: Costs, download limits
- **Insecure**: encryption, authentication, proxy
- **Zero touch**: local network config override, last resort
- **Regulations**: radio silence mode
- Managed by the **NIM** microservice (Network Interface Manager)
- **wwan** and **wlan** microservices manage LTE/WiFi adapters

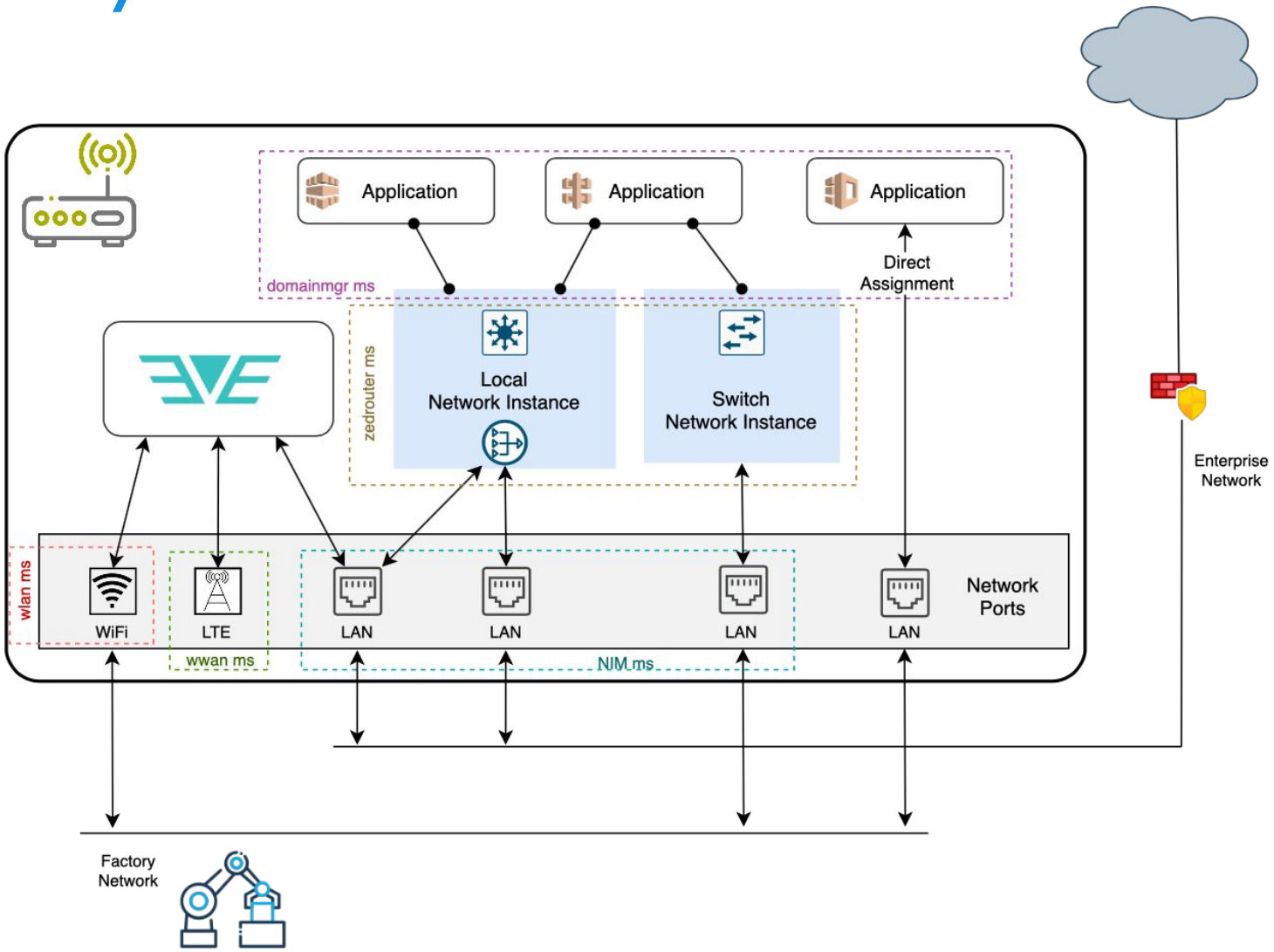
Network Instances

- **Virtual switch** with different forms of external connectivity
  - **Local**: L3, NATed
    - Features: IPAM, DHCP, DNS, ACLs, HTTP Web-Server for metadata
  - **Switch**: L2, bridged with single uplink port (or none)
    - Features: VLANs, ACLs, HTTP Web-Server for metadata
  - **VPN (beta)**: L3, Site-to-Site IPsec VPN
    - IPsec implementation provided by strongSwan/kernel
- Dedicated uplinks vs. overlap with mgmt port
- Uplink testing and failover
- Flow/DNS logging, Interface metrics
- Managed by the **zedrouter** microservice

Application Network Adapters

- **VIFs**
  - Interfaces connecting applications with network instances
  - e1000 (emulation) or virtio-net (paravirtualization)
  - Provide applications with network instance services
  - Data-plane: NIC -> Host kernel -> Guest kernel
- **Direct assignments (PCI passthrough)**
  - Entire NIC assigned to application (until SR-IOV is supported)
  - Uses vfio-pci / xen-pciback driver and IOMMU for safe DMA
  - Bypasses network instances
  - Data-plane: NIC -> Guest kernel
- Both Managed by the **domainmgr** microservice

Visual Summary



EVE Networking
Work in Progress

THE LINUX FOUNDATION

VLAN and LAG Support

- **VLANs**, **LAGs** and **VLANs over LAGs**
  - for EVE mgmt and local network instances
  - VLAN sub-interfaces and bond interfaces
- Currently only switch NIs support VLANs
  - VLAN endpoints are inside apps, EVE applies BD VLAN filtering
- Use-cases:
  - (VLANs) **Isolate** mgmt traffic from application traffic
    - Different ACLs, separate traffic shaping/policing, etc.
  - (LAGs) Interface **load-balancing**, **Failover**
- Various bonding modes (from Linux bonding driver)

Hardening & Refactoring

- Use VRFs to **reinforce isolation** between network instances
- NIM and zedrouter refactoring:
  - Improve **abstraction layering** and **separation between concerns**
  - Limited interleaving to **simplify** the state machine
  - **Replaceable** and **testable** components
- Graph theory based configuration processing
  - Represent **intended config** and dependencies with **graph**
  - **Generic** reconciliation algorithm (state diff and operation ordering)
  - Easily **extensible** (new feature = new nodes and edges)
- Logging improvements
  - Take full advantage of **semantic logging**
  - Log transitions of important state variables

EVE Networking
R&D for 2022 (NFV)

THE LINUX FOUNDATION

Network Performance Optimizations

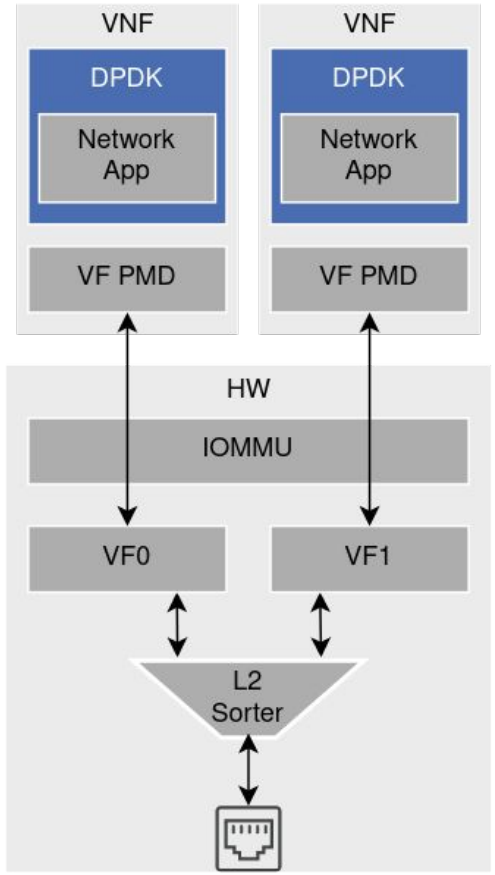
- Target: using EVE OS as **NFV platform**
- Performance optimization techniques:
  - **Context-switch reduction** (e.g vhost-net)
  - **Eliminating CPU interrupts** (e.g DPDK PMD)
  - **Kernel/Host bypass**
    - Avoid hypervisor overhead
  - **HW Offload**
    - Reduce CPU utilization and cache thrashing
  - **Hugepages**
    - Avoid TLB misses
  - **Fast-Slow path separation**
    - Control traffic takes the slow path
    - Majority of data traffic takes the fast path

SR-IOV Support

- Allows a single PCI Express device to present itself as several virtual NICs
- **Physical Function (PF)**
  - Fully featured PCIe functions
  - Allows to control and configure the device
  - Typically assigned to the host
- **Virtual Function (VF)**
  - Lightweight PCIe function capable of data transfer only
  - Typically assigned to a VM
- SR-IOV NICs contains an **embedded switch** (aka L2 sorter)
  - Packet forwarding based on dst MAC or VLAN ID
- **Pros**: efficient use of resources, hypervisor bypass, offloaded switching

DPDK Integration

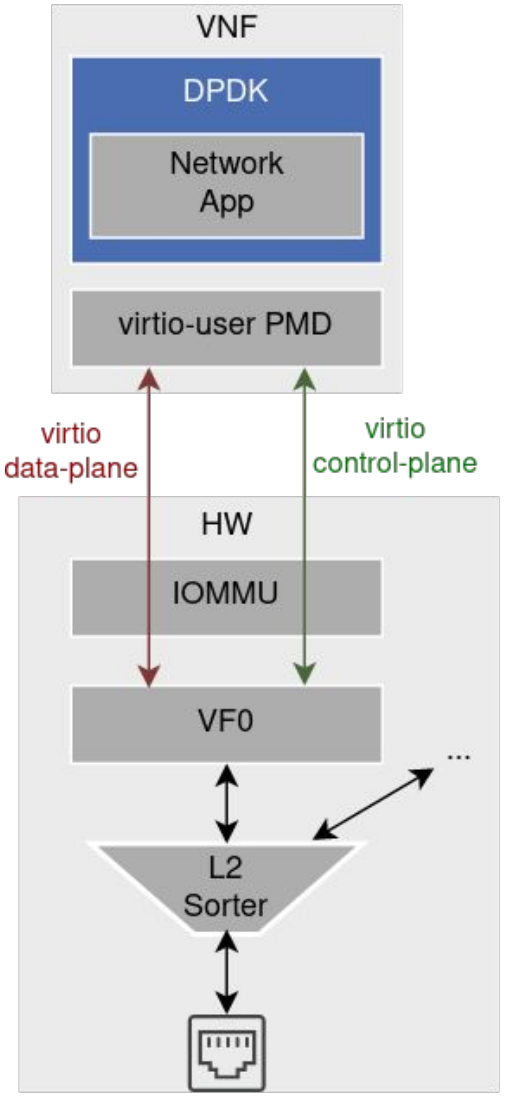
- Framework for fast packet processing in data-plane applications
- **Kernel bypass** to avoid kernel-user context switching
- Uses **Poll Mode Drivers (PMDs)** to avoid interrupts
- Uses **hugepages** to avoid high rate of TLB misses
- PMDs are CPU intensive (cons: wasting cycles)
- **CPU pinning** is a necessity



Example: SR-IOV with dpdk-based apps

SmartNICs

- **Programmable** NICs
- Able to **offload** network functions from server CPUs
- 3 Different approaches: ASICs, SOCs, FPGAs
  - Differ in programmability, performance and cost
- Able to offload an entire vswitch (e.g. OVS)
- Custom data-plane can be implemented in C or P4
- SmartNIC can be used for **fast-path**
- EVE could offload ACLs, NAT, IPsec, etc.
- **virtio offloading** for hardware-agnostic NFVs
- VNF offloading (EVE as mediator vs. passthrough)



Example: SmartNIC with virtio offloading

Thank you!

THE LINUX FOUNDATION