Taking Control of your SmartNIC

Andy Gospodarek (Broadcom)
Or Gerlitz (Mellanox)

What is a SmartNIC?

What a SmartNIC is NOT

- Not a NIC with **only fixed function hardware** capable of stateless or stateful offloads (VLAN, Tunnel, Flow, etc)
- Not a NIC that has hardware with **programmable datapath written using a specific language** (eBPF, P4, HDL, etc)
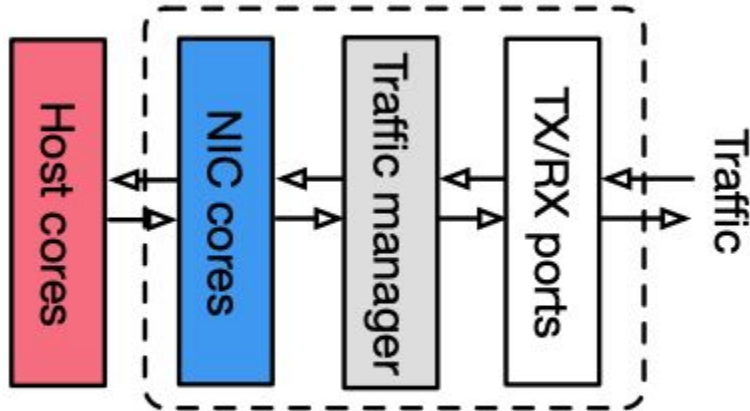
What a SmartNIC is

- A SmartNIC allows a server operator to move control plane applications from server directly to general purpose cores on the NIC (SoC)
- Using definitions from [IETF 105: How NICs work today](#), a SmartNIC is a ***Programmable NIC with General Purpose Processor*** running Linux
- In addition to offloading control plane to general purpose cores, the dataplane might also be offloaded to fixed function hardware (ASIC)
- When needed, further per packet processing such as crypto / compress / pattern matching applied using Linux frameworks on the NIC cores or further offloaded to HW accelerators
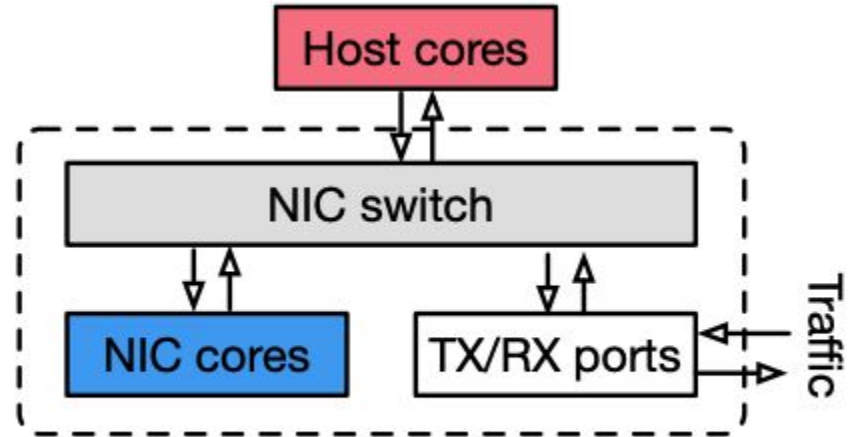
SmartNIC Architecture
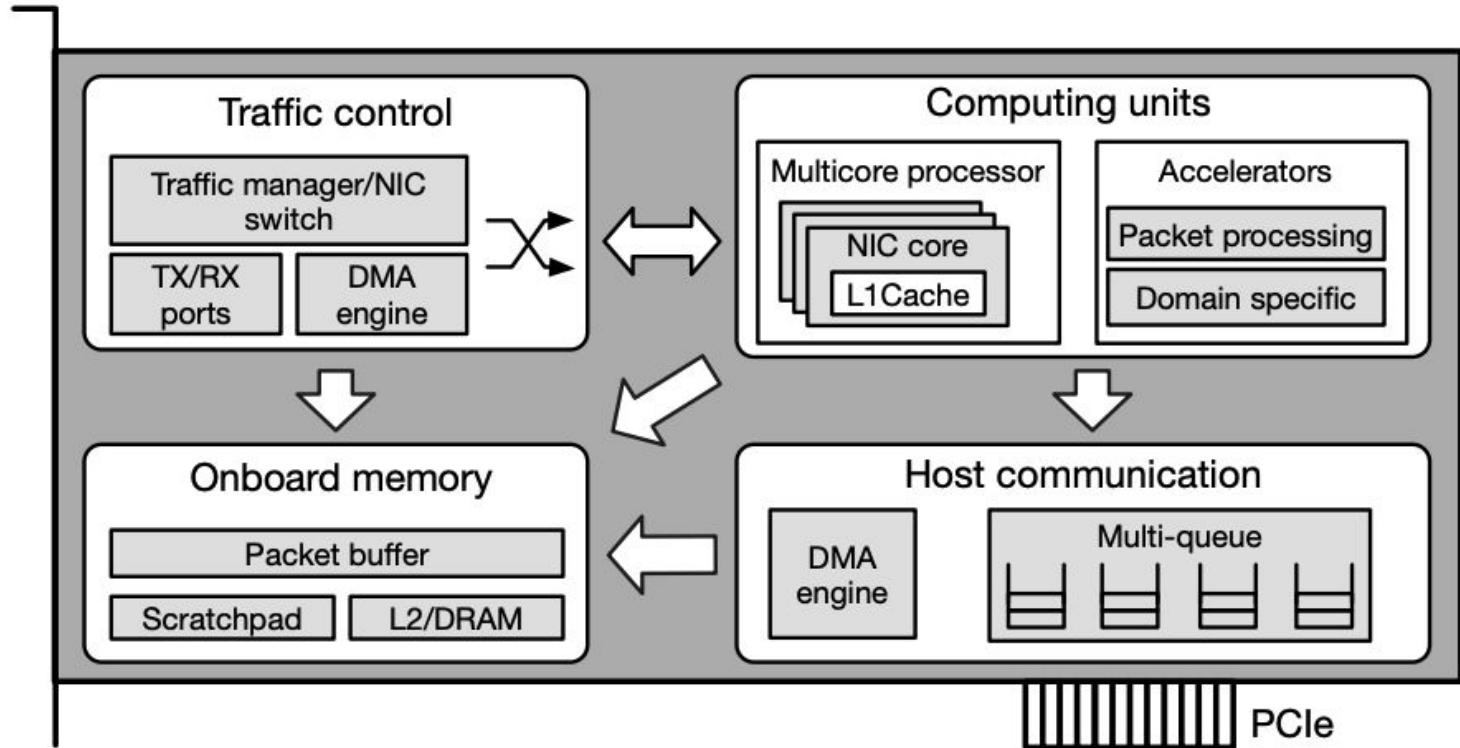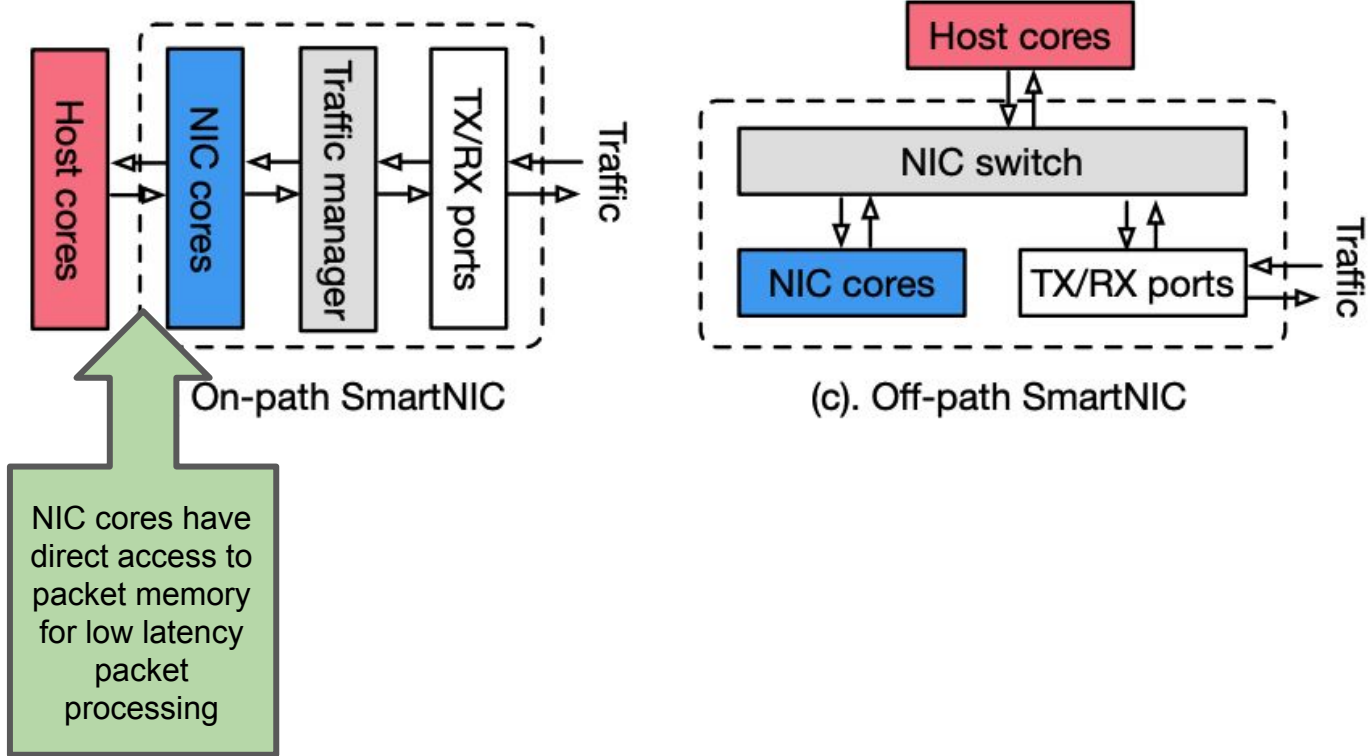
On-Path (b) vs Off-Path (c) Architecture
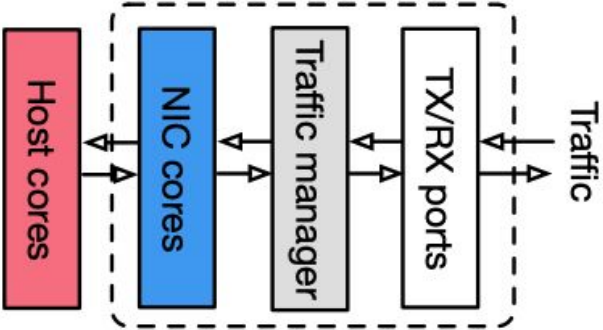

(b). On-path SmartNIC

(c). Off-path SmartNIC

Benefits of On-Path Architecture

On-path SmartNIC: Host cores → NIC cores → Traffic manager → TX/RX ports ← Traffic

NIC cores have direct access to packet memory for low latency packet processing

(c). Off-path SmartNIC: Host cores → NIC switch → NIC cores, TX/RX ports ← Traffic
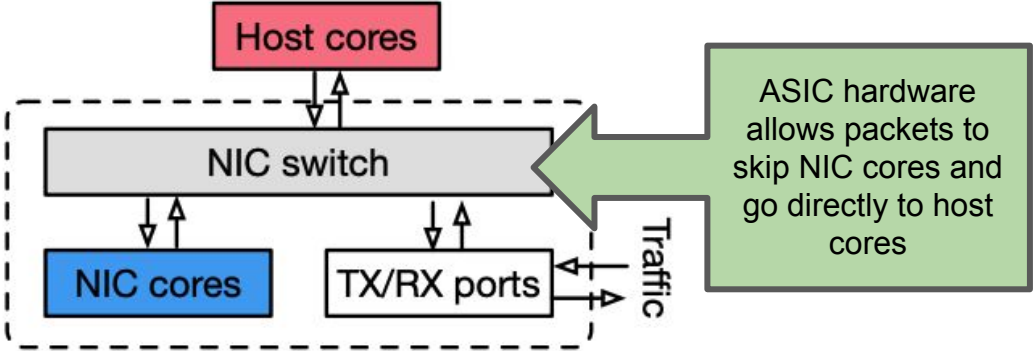
Original Image: https://homes.cs.washington.edu/~mgliu/papers/iPipe-sigcomm19.pdf

Benefits of Off-Path Architecture
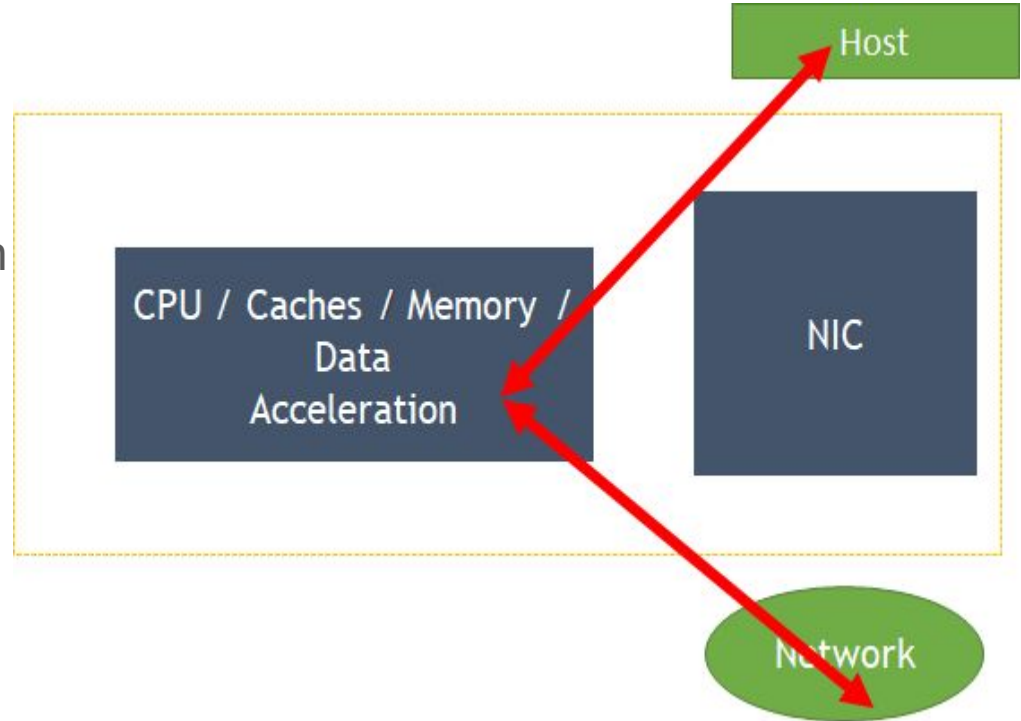

(b). On-path SmartNIC

(c). Off-path SmartNIC

ASIC hardware allows packets to skip NIC cores and go directly to host cores

On-path vs Off-path -- Why does it matter?

- On-path SmartNICs typically expose a low-level programming API
    - Typically uses SDK or other low-level programming interface to access hardware blocks on the SmartNIC.
    - Different vendors would likely require a different software implementation
- Off-path SmartNICs are more programmable or flexible
    - Typically the same dataplane drivers (kernel, DPDK, etc) are used on host and SmartNIC
    - Hardware vendor or 3rd parties can provide an OS to give users control over their own destiny
- **Focus of this talk will be Off-path SmartNICs**
    - **Broadcom Stingray**
   - **Mellanox BlueField**

Off-Path SmartNIC Store and Forward Model

- Run any packet processing application (control and dataplane) on SmartNIC
- Examples include Open vSwitch (software-only), VPP, eBPF/XDP/AF_XDP programs, or other custom applications
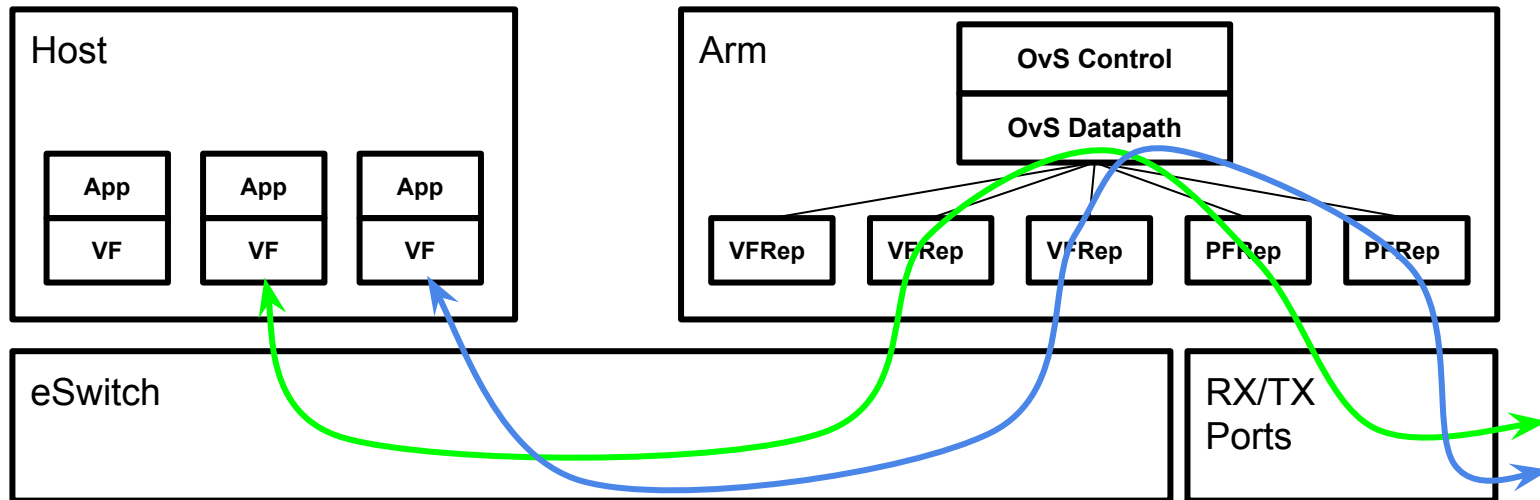- Performance limited by SmartNIC core speed and memory bandwidth
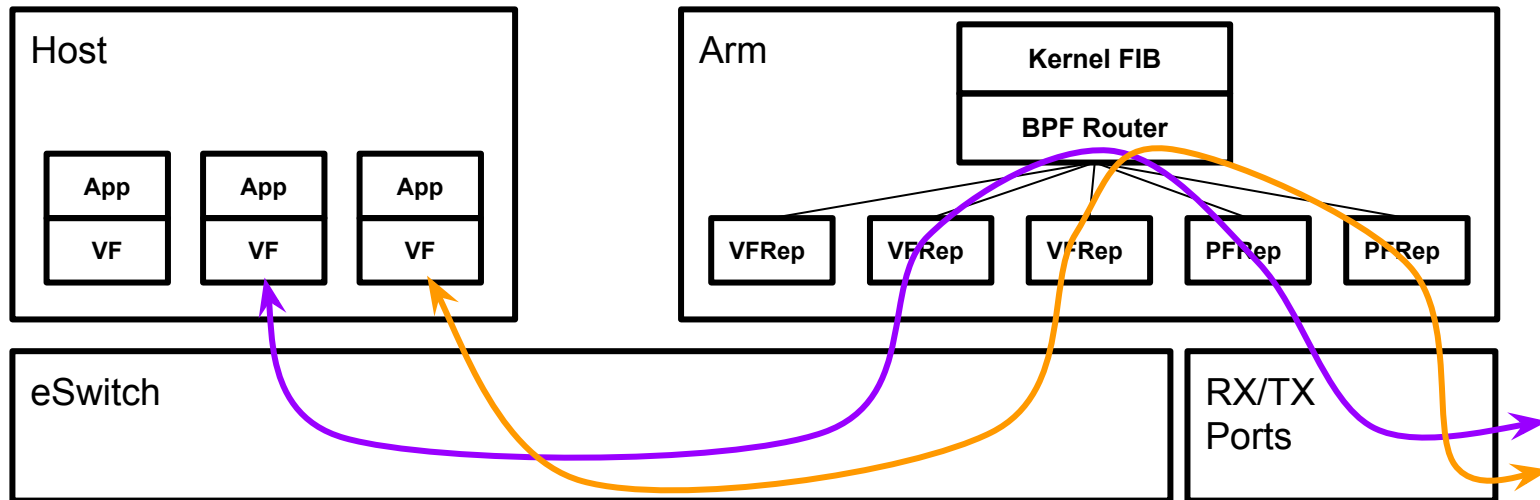
Off-Path SmartNIC Inline Processing Model

- Packets go directly from network to host. No additional copy to SmartNIC cores.
- Control-plane applications can offload operations to hardware:
  - match/action
  - encryption
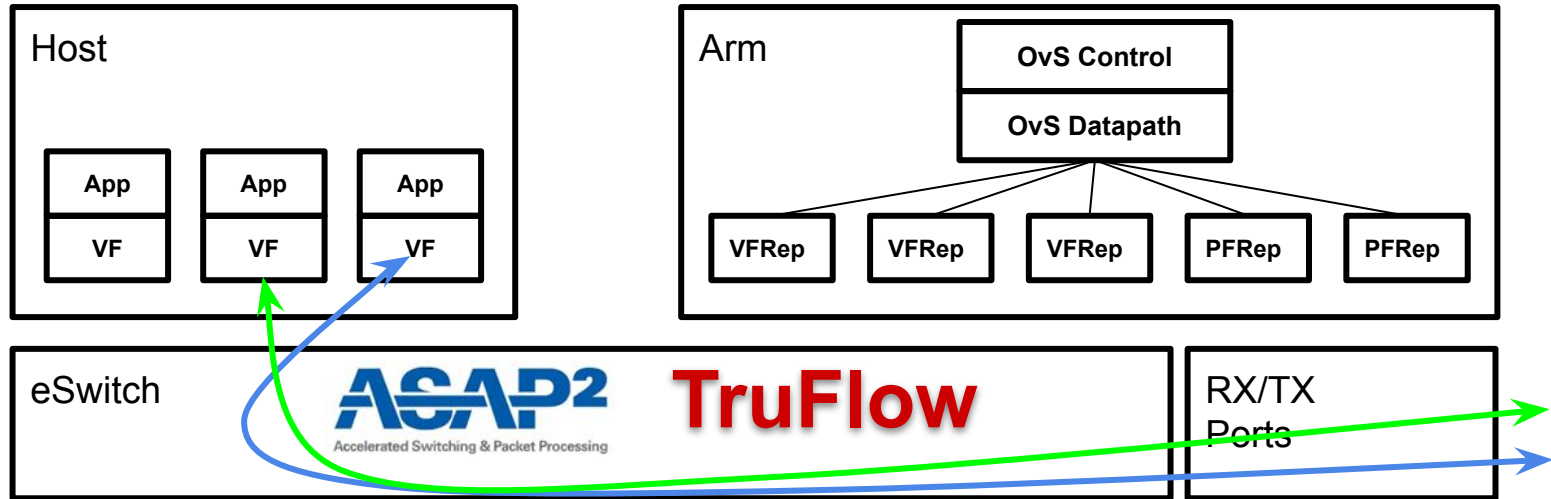 - compression
  - regular expressions
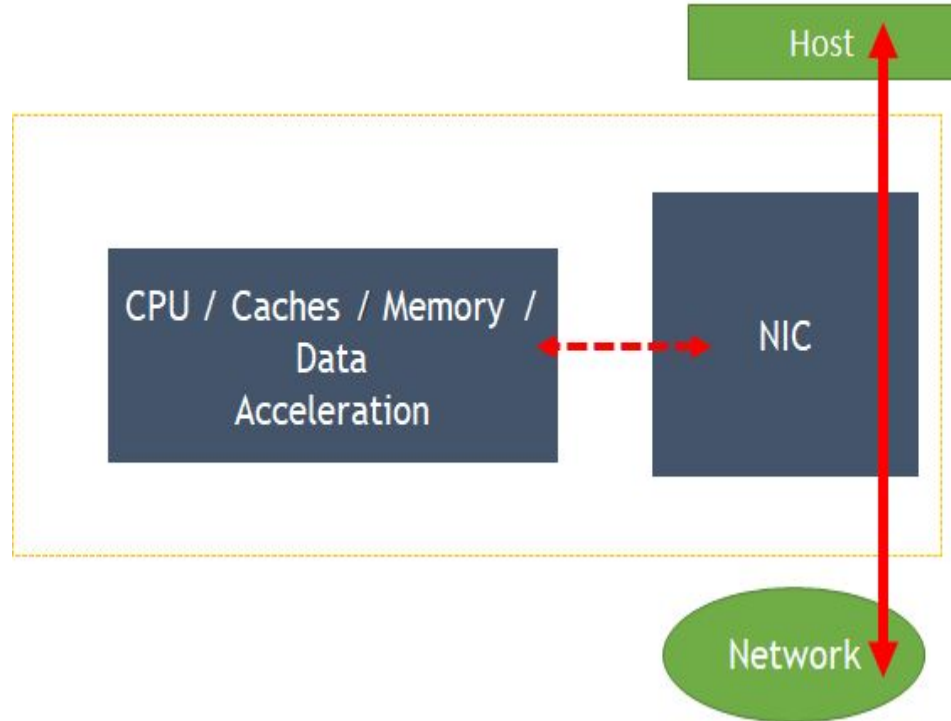
SmartNIC Store and Forward

SmartNIC Store and Forward

SmartNIC Inline - OVS offload

SmartNIC Inline FOSS Ecosystem

- Variety of vendor independent implementations exist to offload network datapath. Identical APIs and drivers used on SmartNIC and Host. Software can transition easily from running on Host to running on SmartNIC
    - Linux kernel data-path via TC/flower and NFT (Connection Tracking)
    - DPDK data-path via RTE flow

- Further HW accelerators introduced on SmartNICs and exposed through Linux kernel APIs enable larger coverage for inline use-cases kTLS and IPSec future kernel APIs for compression and regex

OpenStack SmartNIC Integration

- In legacy installations compute (Nova) and networking (Neutron) provisioning agents run on the host
- In SmartNIC installations networking (Neutron) should move to SmartNIC cores, but compute (Nova) stays on server.
  - More testing needed to validate the communication between all components reliable enough to separate components into different CPU complexes in more cases
- Users have deployed SmartNICs with OpenStack successfully, but currently no turnkey OpenStack solution 🙁

Neutron with `networking-ovn`

- OVN's distributed architecture means that `networking-ovn` may overtake `networking-ovs` as the primary Neutron plug-in for OpenStack deployments
- `networking-ovn` plugin configures OVN Northbound DB
- Configuration flows through OVN Southbound DB and `ovn-controller` agent runs directly on SmartNIC and program datapath
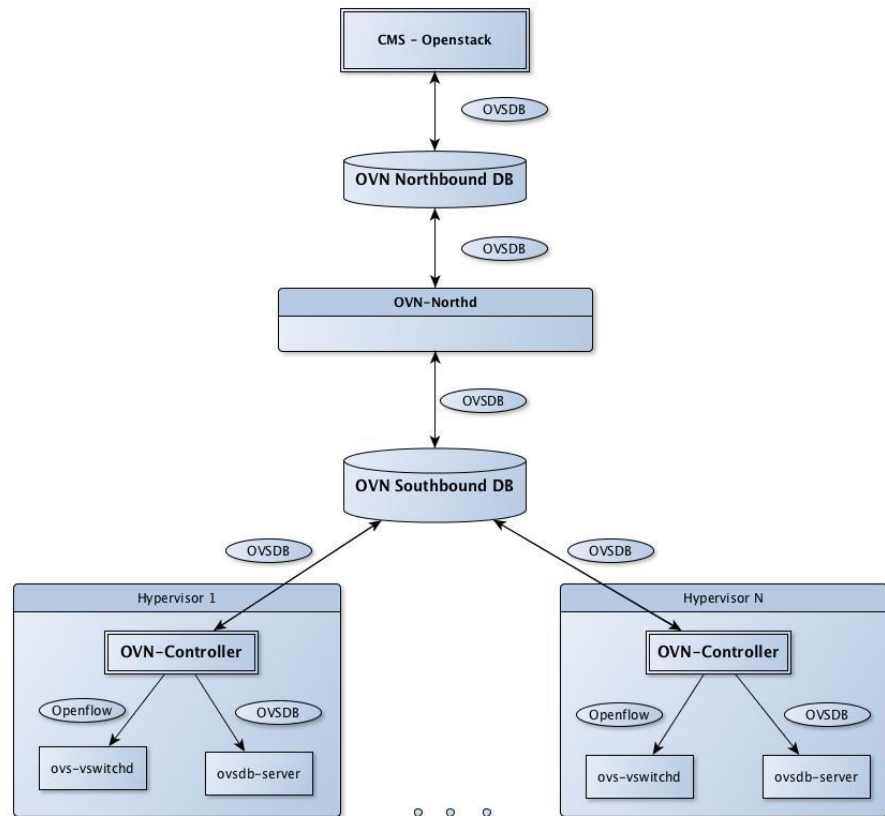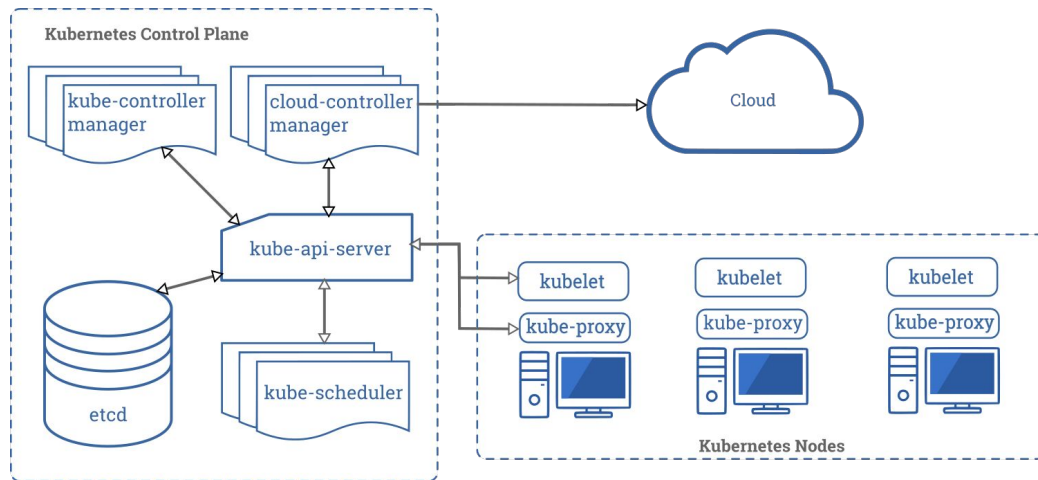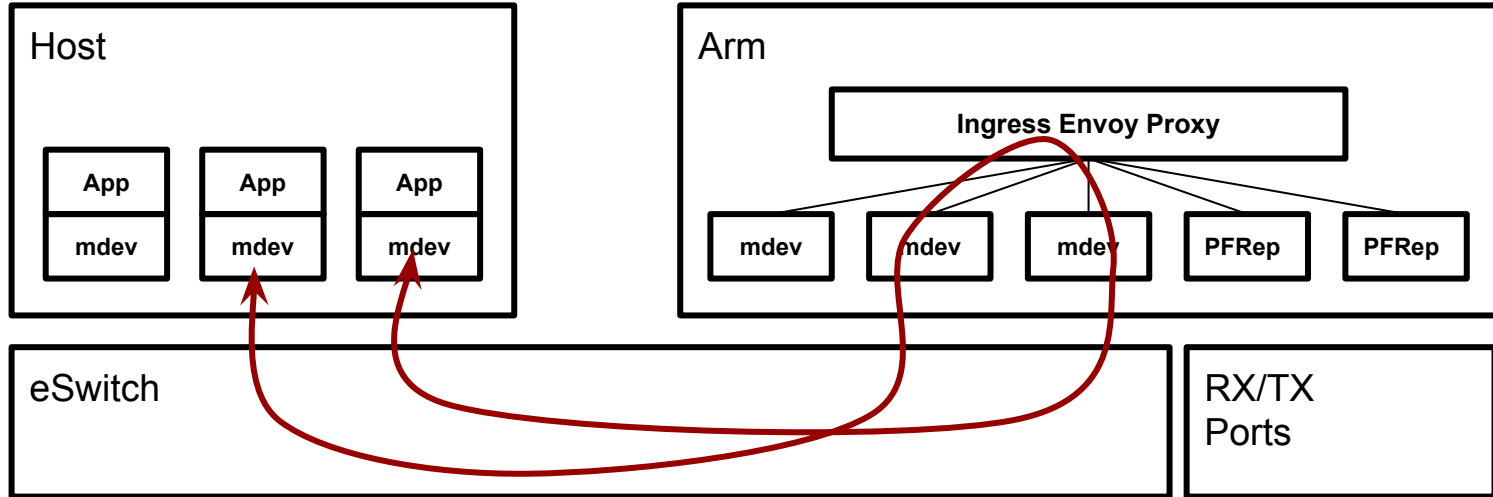
Kubernetes

Decoupled architecture might make k8s a better fit for SmartNICs

`kube-proxy` could be moved from Node to SmartNIC to offload datapath processing

Another good candidate would be to move Ingress Controller to SmartNIC

# Envoy on Stingray SmartNIC (Store and Forward)

Testing with Stingray SmartNIC reveals that communication between applications via Envoy on SmartNIC provides useful offload. Stingray can offload ~6 Host cores normally spending time performing packet processing.

Independent Management of a SmartNIC

- In true Open Source fashion some will want to manage their SmartNIC as an independent networking element
- Traditional tools may be used for management and configuration rather than larger integrated frameworks
- This is particularly popular in rising trend of Baremetal Server deployments
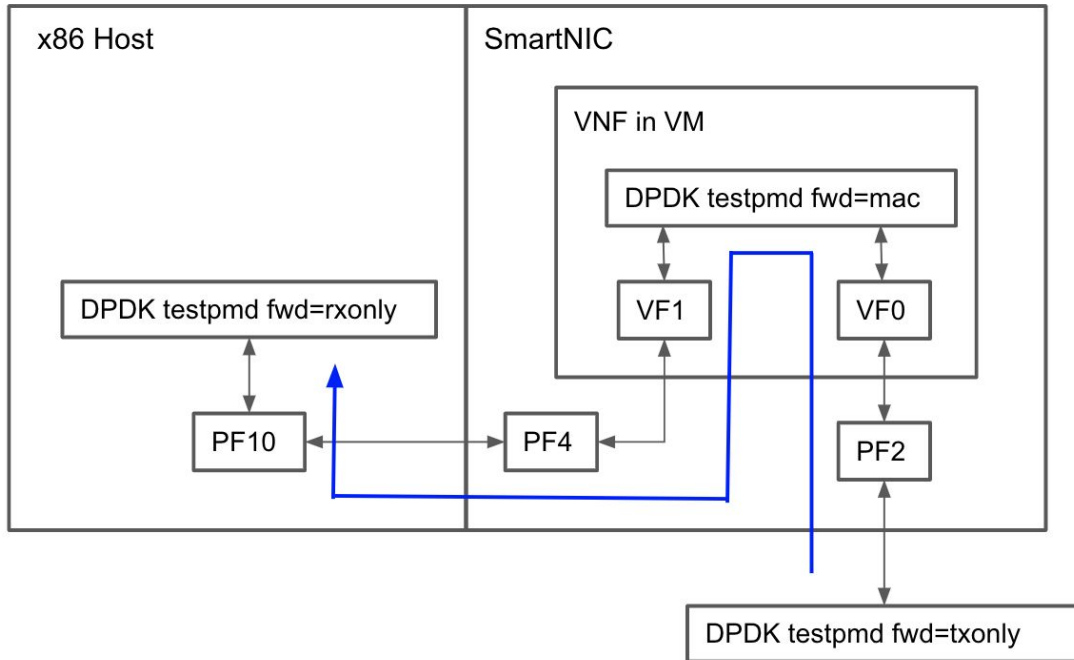
Using SmartNIC to run a VNF
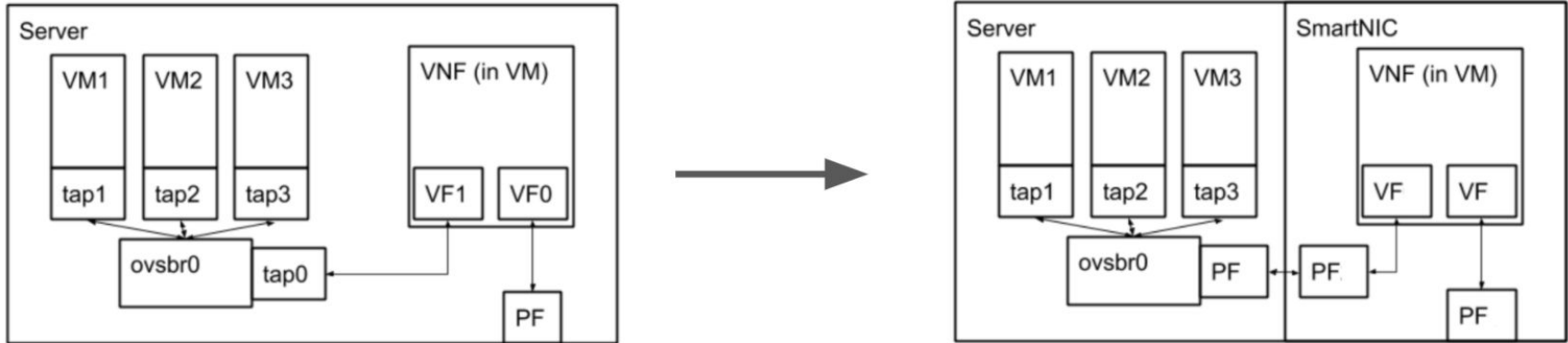
Allows VNF providers to create images that can also be run directly on SmartNIC.

Broadcom Stingray running VNF

Example and scripts for running on Broadcom Stingray available on [GitHub](#)
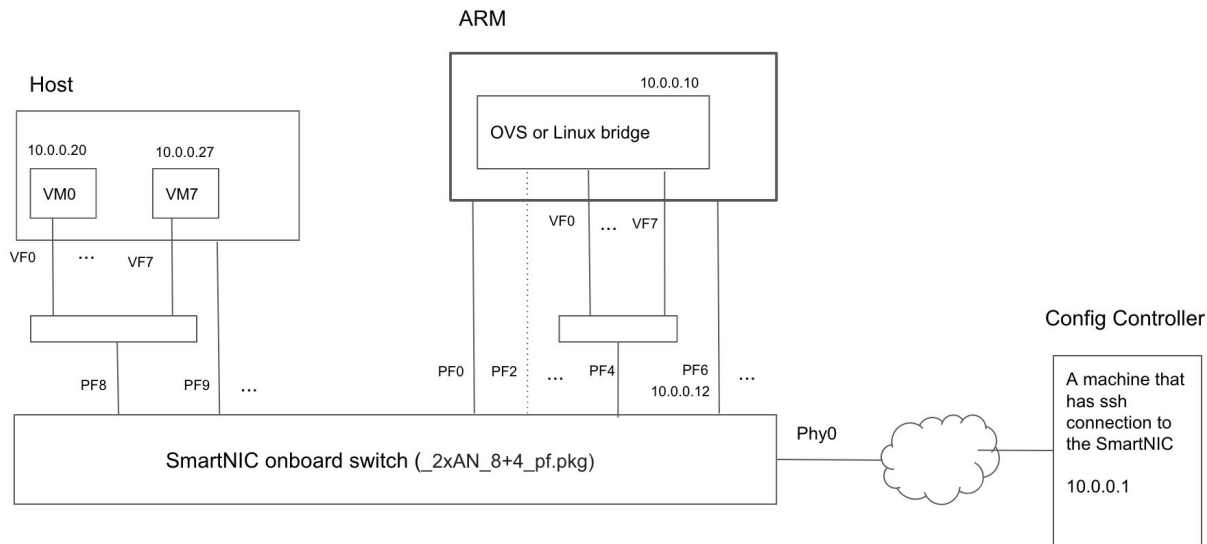
Framework and instructions available for creating custom VNF based on Ubuntu Cloud Images

SmartNIC Configuration with Ansible

Since Linux is running on SmartNIC cores, standard tools like Ansible can be used for configuration.

A full example of how to do this with Broadcom Stingray available on GitHub.

SmartNIC in Baremetal Clouds

- Idea: use SmartNIC as the means to leapfrog latest networking and storage technologies into legacy servers and appliances
- Server operator need not have latest technology/vendor driver
- Networking: SmartNIC exposes common NIC HW function to server
  - virtio - has spec and wide deployment
- Storage: SmartNIC exposes common disk HW function to server
  - NVMe PCI - has driver on all OSs, can boot from, etc
  - no need for local storage or NVMe fabrics driver on the host
- FOSS facilities for the SmartNIC operator to integrate with
  - Virtio: kernel and DPDK vDPA infrastructure
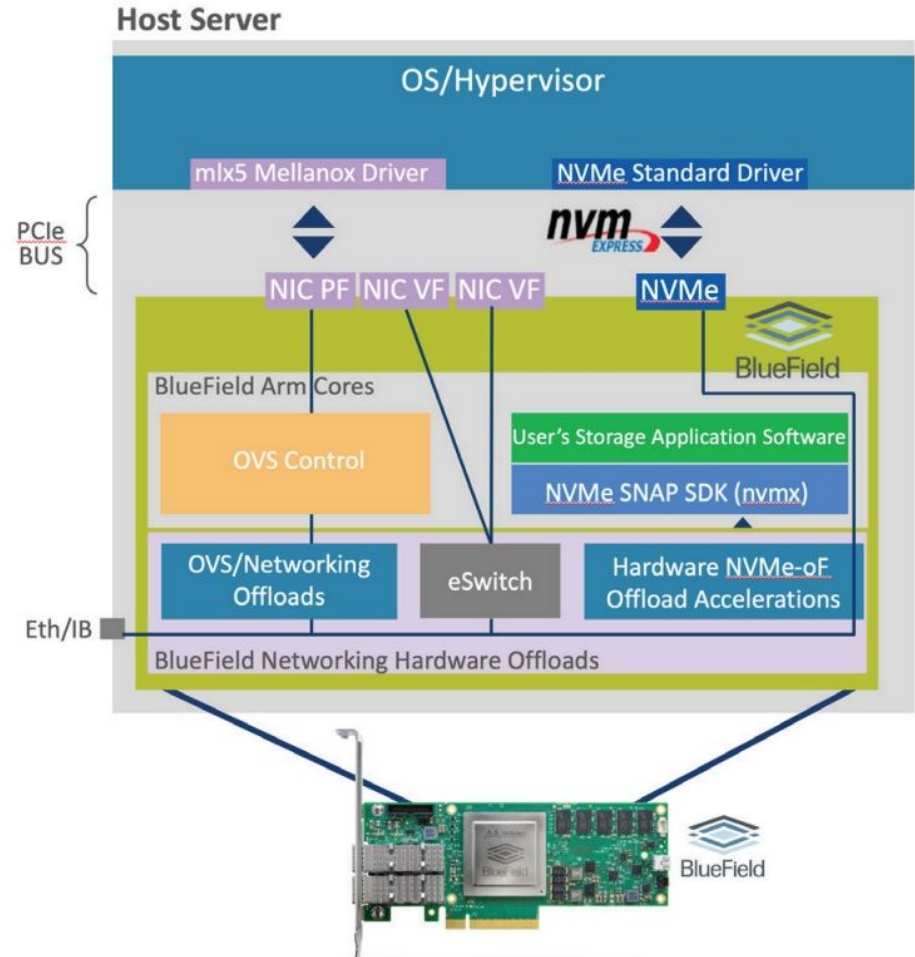 - Storage: kernel and SPDK storage stack

BlueField SNAP

Storage-Defined Networking Accelerated Processing

Framework to emulate NVMe local storage, connecting to remote storage.

Directly (inline) without going through SmartNIC Cores (NVMe/RDMA only).

For different transport protocols (e.g NVMe/TCP) use Store and Forward proxy.

Key Takeaways

- Open Programmable NICs need to be simple to use. Must be capable of running Open Source Operating systems and applications
    - Use of standard automation tools a requirement
    - Integration into application suites like OpenStack or K8s aids adoption of SmartNICs but the distributed nature of those frameworks does not make full integration a requirement
- Two major differences between use cases
    - Virtualization/Host Controlled -- server operator controls the SmartNIC
    - Baremetal/Remote Controlled -- network operator controls the smartnic
- Baremetal provisioning direction is promising for users