arm

Empowering Edge Software Stack on Arm with High Performance Container Networking Technology

Trevor Tao, Song Zhu

Mar/2020

Agenda

- Introduction
- Arm Edge Software Stack Architecture
- High Performance CNIs(Calico, Cilium…) for Arm Edge Software Stack
- SRIOV and SmarNIC support for Container Networking
- Performance Tests
- Future Work(Provisional)

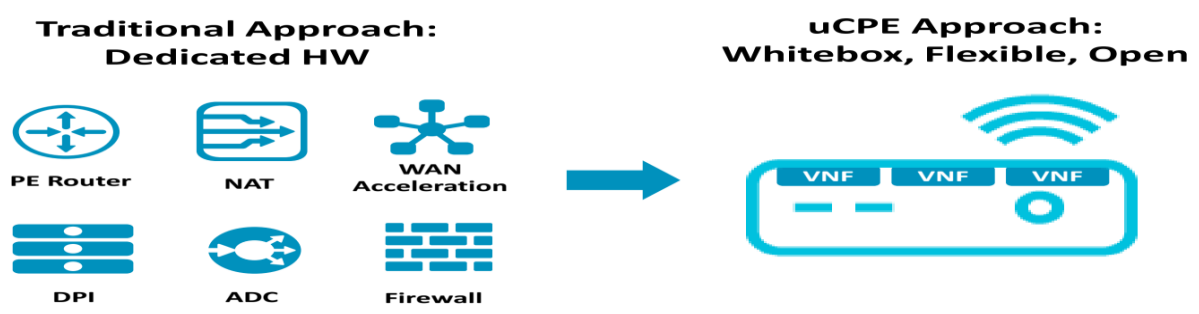
Introduction

arm

Introduction

- The purpose of edge computing and MEC is to bring real-time, high-bandwidth, low-latency access to latency-dependent applications, distributed at the edge of the network. Arm CPUs are cost effective, consume low power and are customizable; making them a preferred choice for edge cloud service vendors.
- On the other side, a high performance, flexible and easy deployable container networking of edge software stack is the key to the success of using Arm platform on edge cloud since the widely used VNFs are deployed by container-based orchestration engine, such as Kubernetes, OpenShift or others.

Universal CPE example of Edge compute

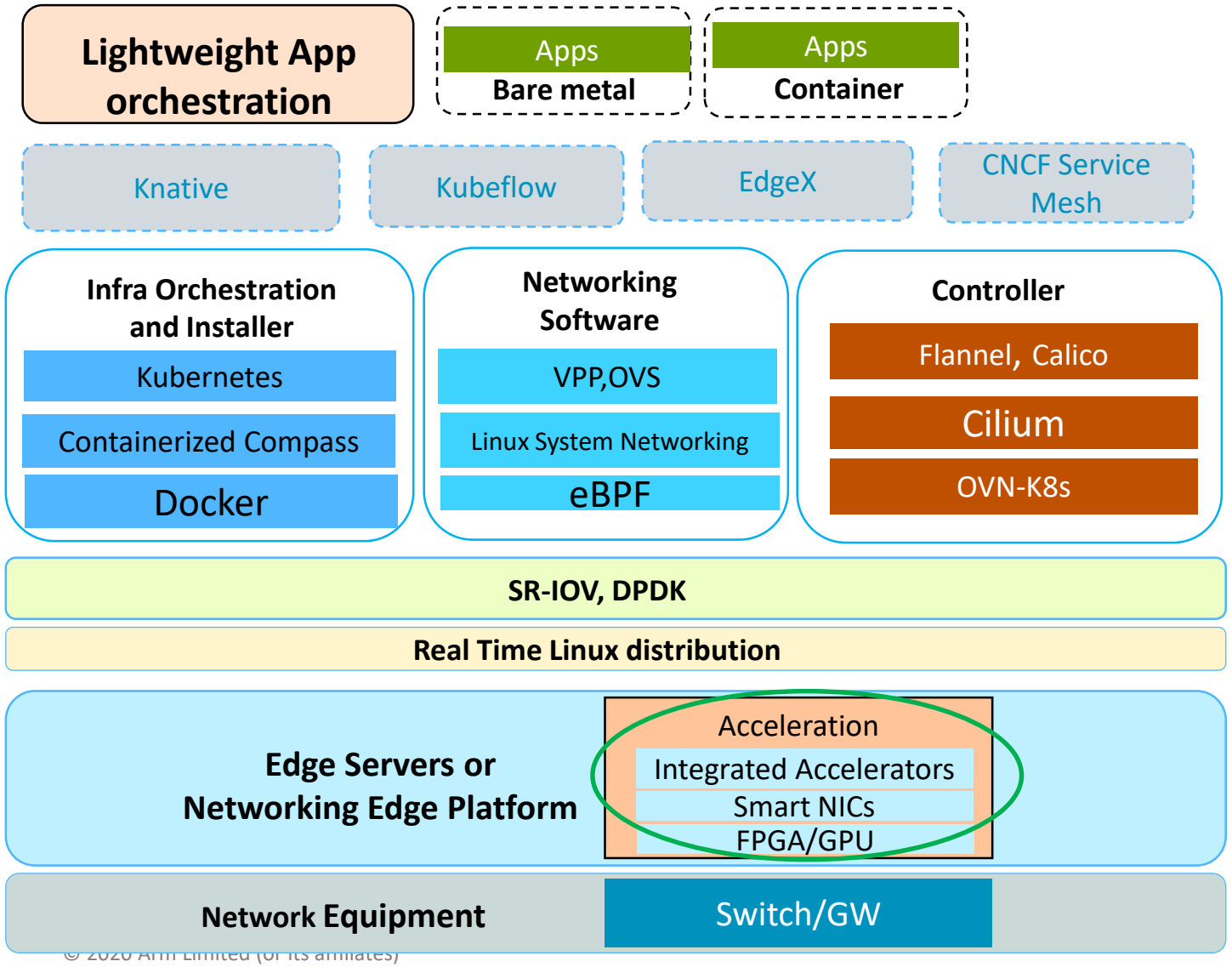


- Service providers looking to simplify customer site deployments
- Replace multiple dedicated appliances with VNFs running on a uCPE – a single, universal platform
  - Customers can provision VNFs from different vendors on uCPE boxes via operator platforms
- uCPE should leverage white box servers with no proprietary extensions

Arm ecosystem leverage price, power, acceleration advantages

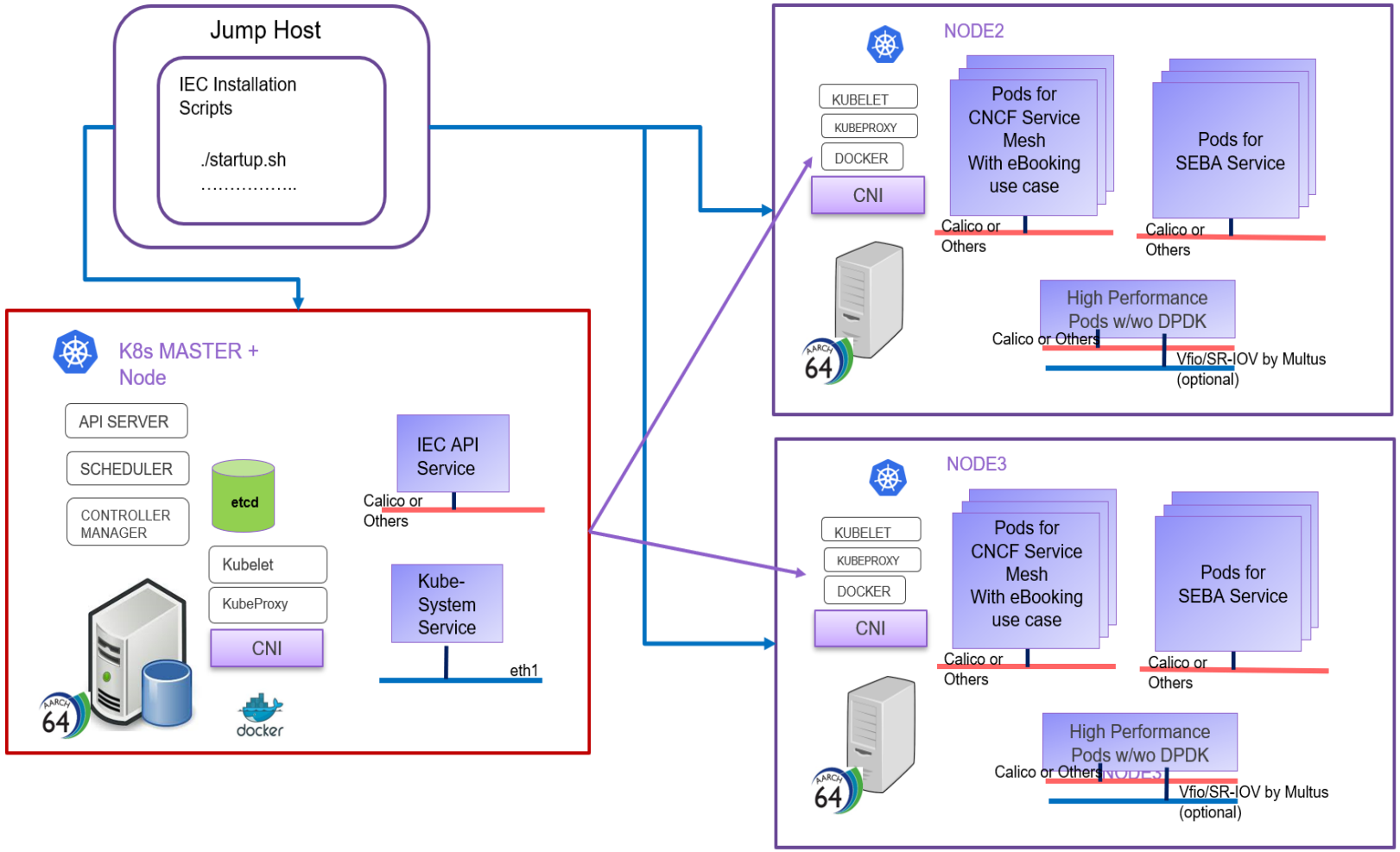
https://community.arm.com/developer/ip-products/processors/b/processors-ip-blog/posts/arm-and-telco-systems-partner-to-redefine-universal-cpe

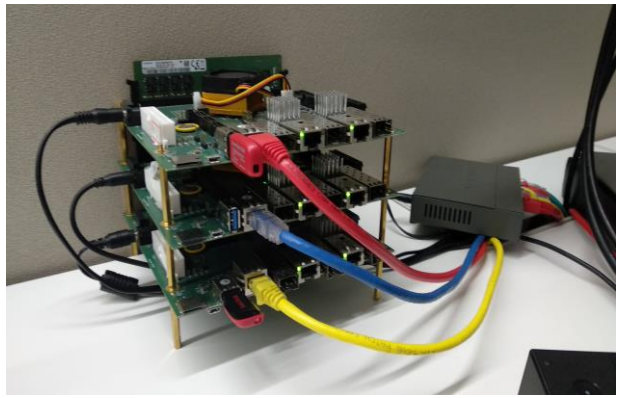
Arm Edge Reference Stack



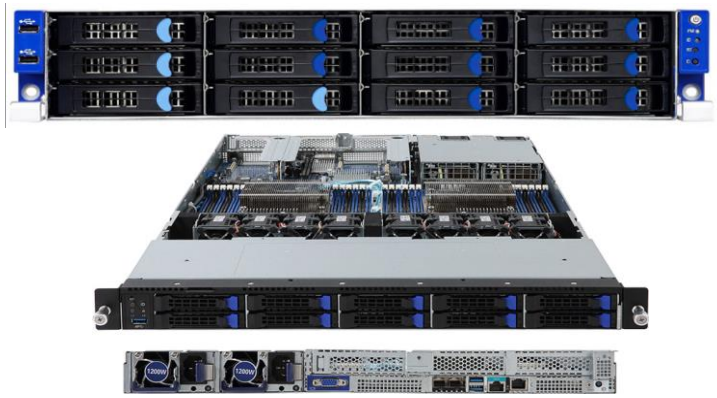
- Heterogeneous Architecture
  - VM, container, bare metal
  - Servers and customized Edge platforms
  - Virtualized NFs and Physical NFs
  - Accelerator interface
- Resource constraints
  - Kubernetes
  - SDN Controller for K8s
- HW Accelerations
  - Integrated accelerators
  - PCIe/CCIX attached accelerator (Smart NICs…)

Arm Edge Stack Deployment Reference Cluster





Small deployment with Marvell Mcbin DBShot Boards



8

Arm server samples used in edge cloud

High Performance Container Networking for Arm Edge Software Stack

arm

High Performance CNIs available for Arm Edge Stack

Things now available in Akraino IEC Arm edge stack as a ref:

IEC Arm Edge Stack

Calico

- pure IP networking fabric
- high-level network policy management by iptables
- Good scalability
- Support direct(non-overlay) and overlay(IPINIP, VxLAN) network connection
- Easy deployment
- Calico-VPP appears

Cilium

- Linux-Native, API-Aware Networking and Security for Containers
- Linux eBPF based network policy, load balance and security which is believed to be with incredible performance
- L3 networking between hosts
- Good scalability too

Contiv-VPP

- uses FD.io VPP to provide network connectivity between PODs
- Native DPDK interface support for phy NIC
- Native VPP ACL/NAT based network policy and access
- Good performance but with rather complex configuration
- Hard to debug

OVN-K8s

- OVS/OVN-controller based K8s networking solution
- Rather good performance with OVS inherited
- Use OVN logical switches/routers to connect Pods and for outside access
- No OVS-DPDK support now

SRIOV

- Direct physical interfaces(**PF/VFs**) support for Pods
- High performance with direct Linux kernel eth driver or DPDK PMD driver
- Usually co-work with other CNIs, such as Flannel, Calico by Multus or other glue CNI
- Need resource description or annotation when do the configuration for CNI and Pod setup

Flannel

- Widely used and almost easiest deployment for a simple K8s networking
- Linux network bridge for pod connection and overlay based communication for inter-hosts access
- Easy to be integrated into other container networking solution, e.g., Cilium
- No good network policy support

Code Review / iec.git / tree

summary | shortlog | log | commit | commitdiff | review | tree
history | HEAD | snapshot

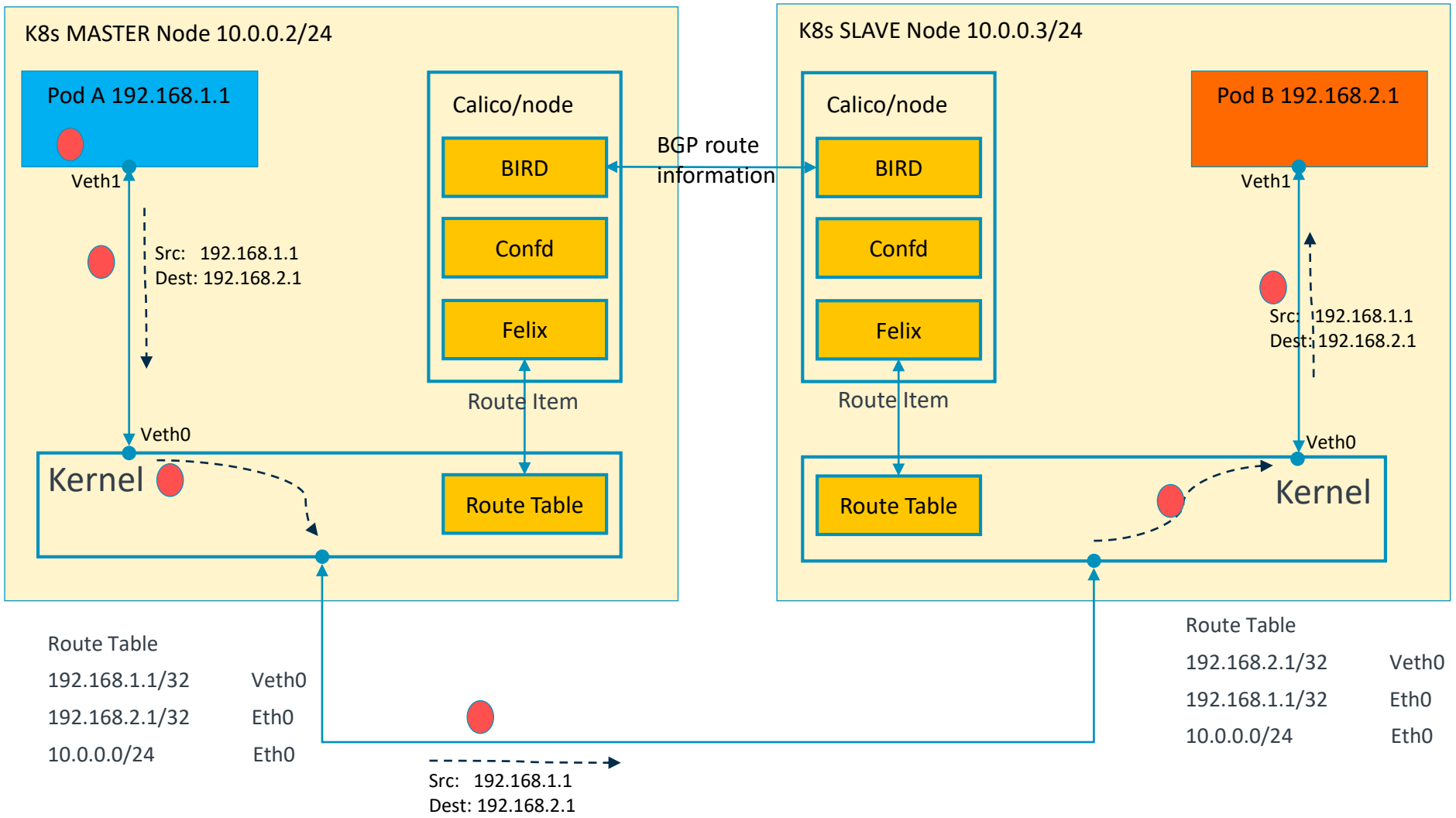
fix issue that deploy cilium on latest image

[iec.git] / src / foundation / scripts / cni /

drwxr-xr-x		..	
drwxr-xr-x	-	calico	tree | history
drwxr-xr-x	-	cilium	tree | history
drwxr-xr-x	-	contivpp	tree | history
drwxr-xr-x	-	danm	tree | history
drwxr-xr-x	-	flannel	tree | history
drwxr-xr-x	-	multus	tree | history
drwxr-xr-x	-	ovn-kubernetes	tree | history

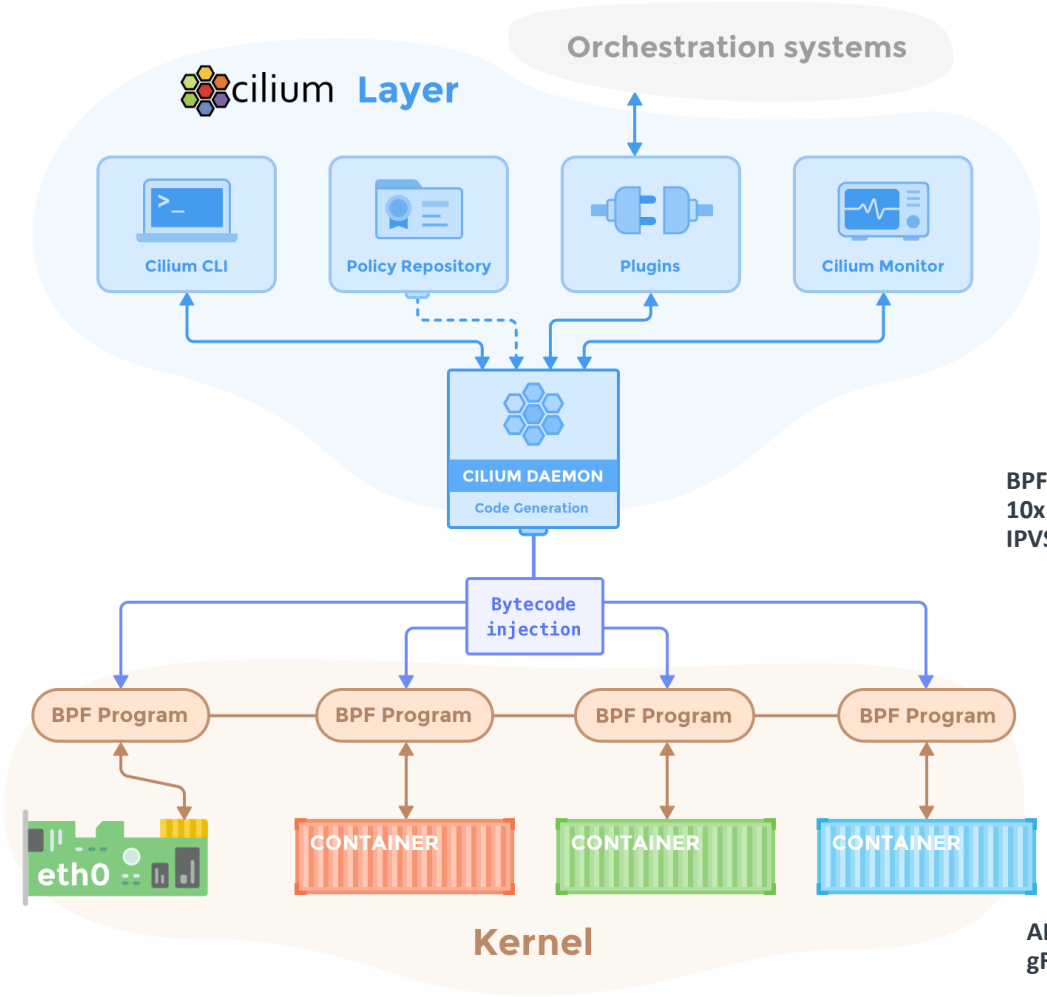
Repo: https://gerrit.akraino.org/r/admin/repos/iec

Calico on Kubernetes



1. Pure IP networking fabric
2. No encapsulation needed when simple L2 connection available for nodes
3. Easy deployment and debug
4. Supporting Kubernetes Network Policy by iptables
5. Good scalability with BGP based routing

Cilium – API Aware Networking



Cilium Architecture

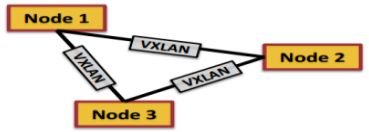
BPF/XDP load balancing 10x performance over IPVS

API Aware(HTTP, Kafka, gRPC)

quick-install.yaml

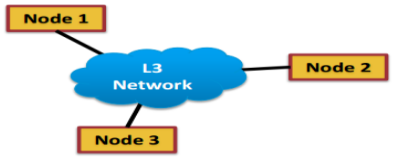
Networking Model: Encapsulation or Direct Routing

Mode I: Encapsulation



No further dependencies

Mode II: Direct Routing

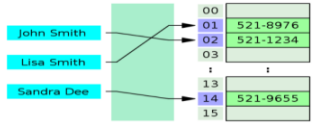


Integrations:
- Cloud routers
- kube-router, BIRD, ...

Kubernetes Services Implementation

BPF-based

- *Per-CPU Hash table*



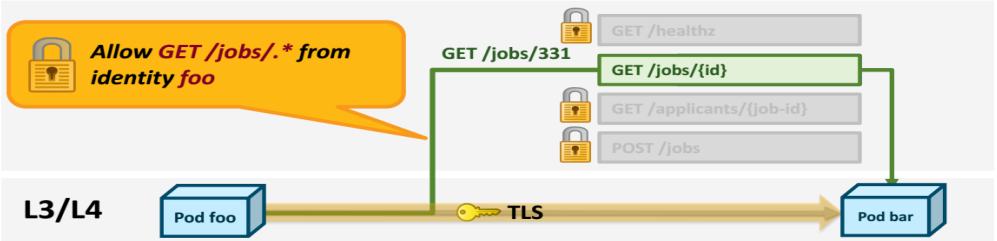
Algorithm	Average	Worst case
Space	O(n)[1]	O(n)
Search	O(1)	O(n)
Insert	O(1)	O(n)
Delete	O(1)	O(n)

iptables kube-proxy

- *Linear List*
- *All rules have to be replaced as a whole*

Algorithm	Average	Worst case
Space	O(n)	O(n)
Search	O(n)	O(n)
Insert	O(1)	O(1)
Delete	O(n)	O(n)

API Aware Security

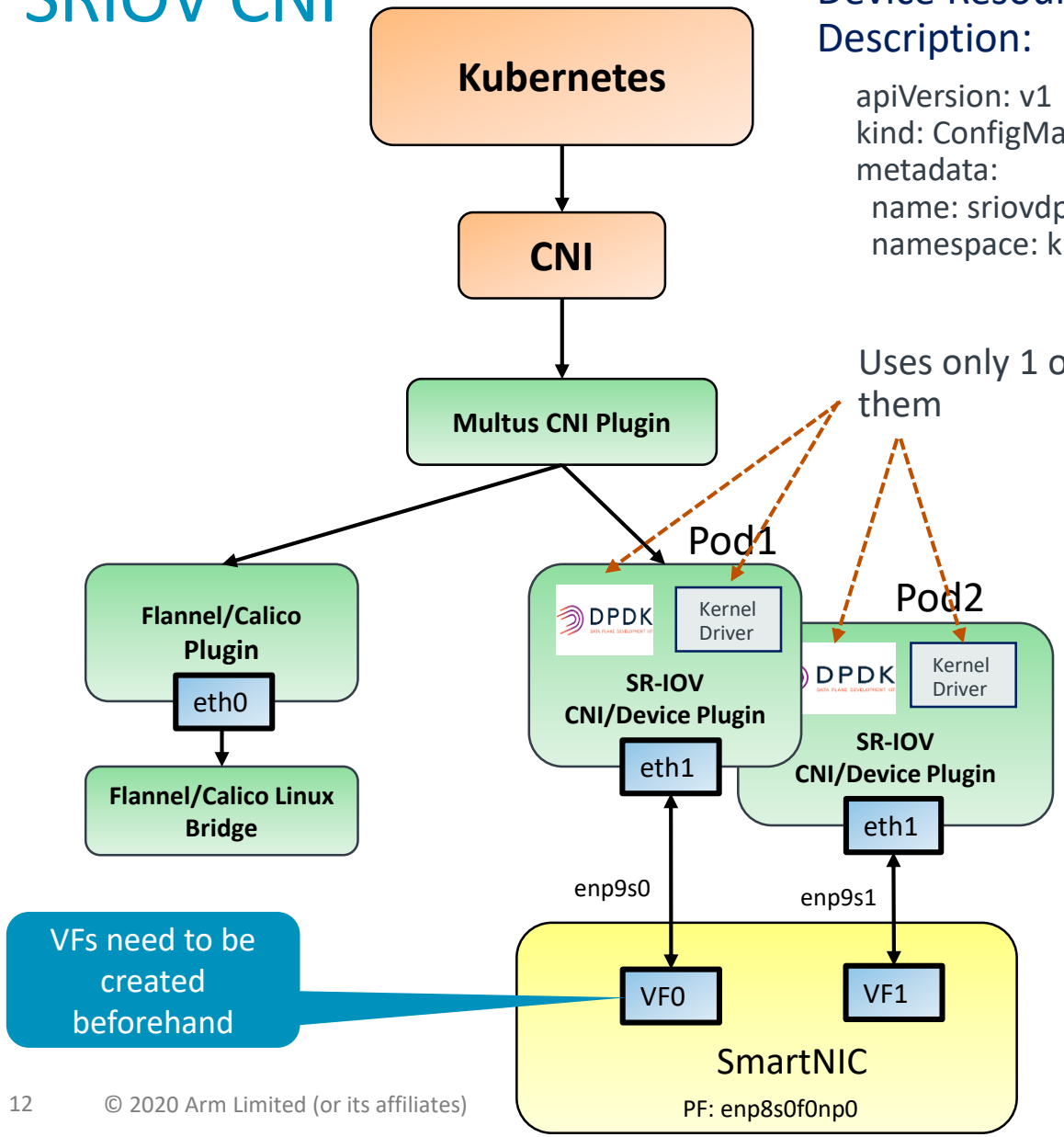


Cilium: API Aware Netowrking & Network Security for Microservices using BPF & XDP

SRIOV and SmarNIC support for Container Networking

arm

High Performance Networking with SmartNIC SRIOV Interfaces by SRIOV CNI

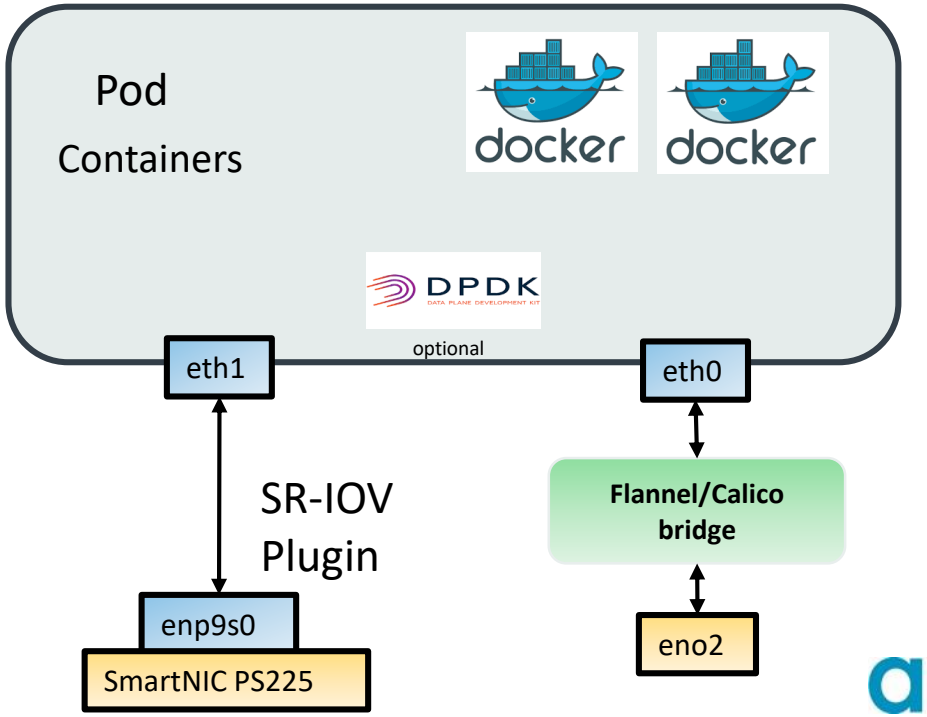


Device Resource Description:

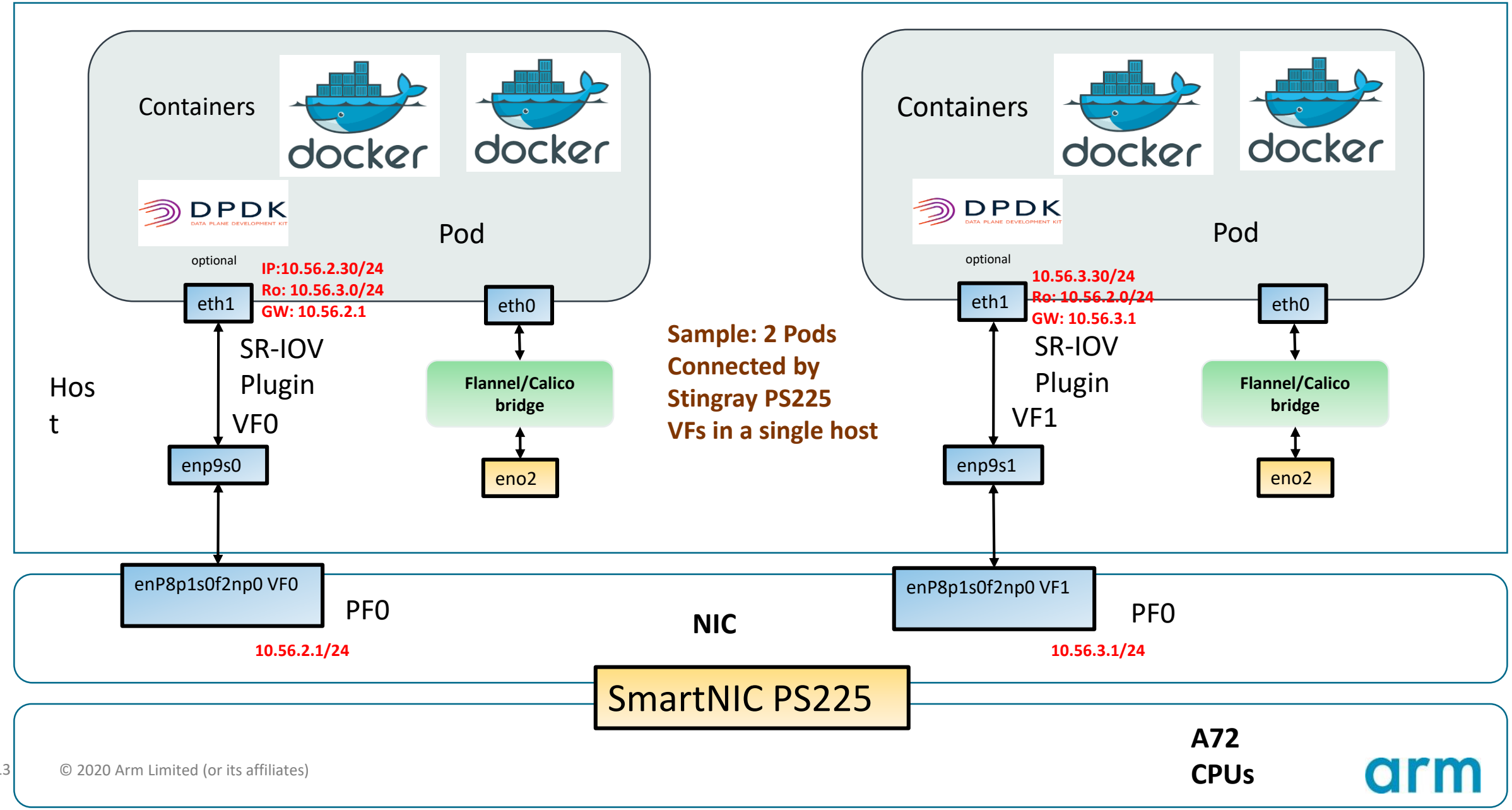
```
apiVersion: v1
kind: ConfigMap
metadata:
  name: sriovdp-config
  namespace: kube-system
data:
  config.json: |
   {
      "resourceList": [{
            "resourceName":
"ps225_sriov_netdevice",
            "selectors": {
              "vendors": ["14e4"],
              "devices": ["d800"],
              "drivers": ["bnxt_en"],
              "pfNames": ["enp8s0f0np0"]
            }
        }, ...
}
      ]
   }
```

For DPDK device driver(another Intel x710 nic as a sample):

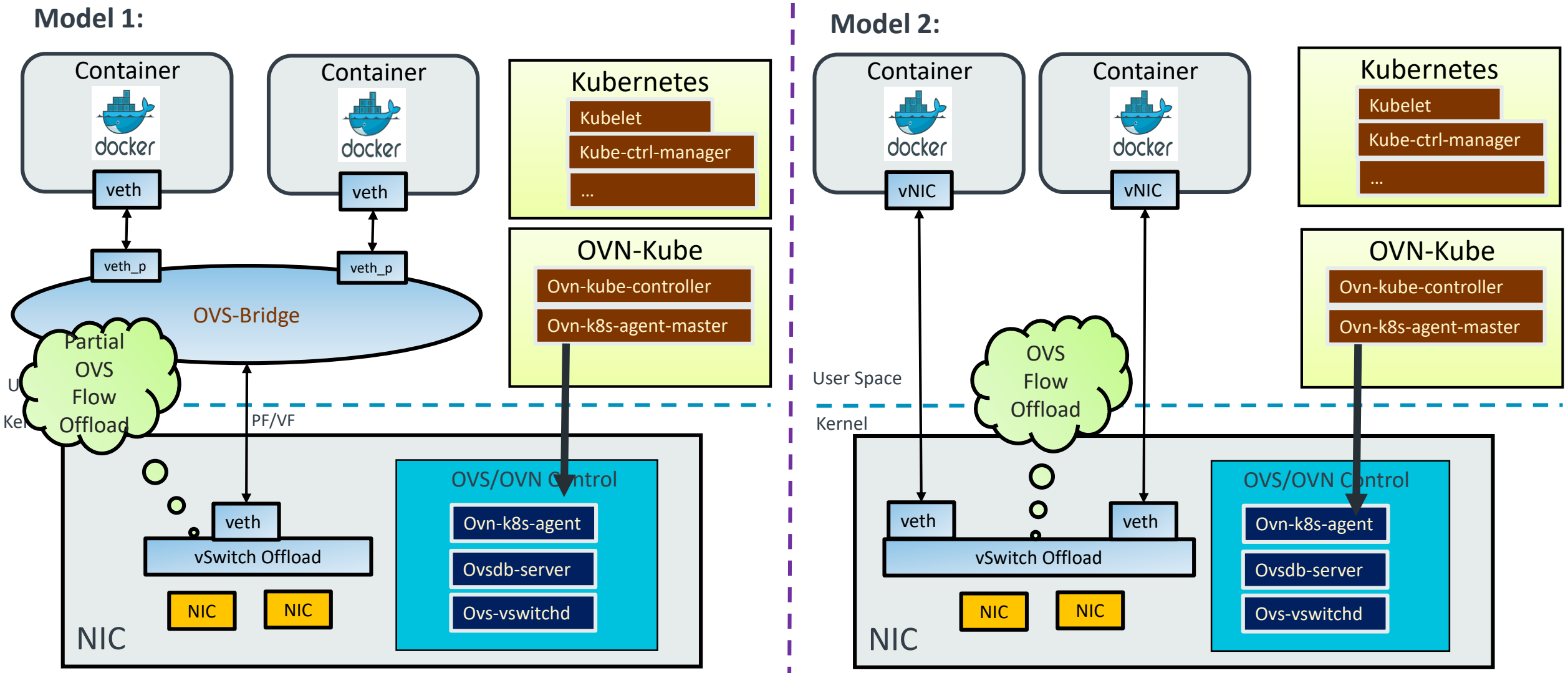
```
{
      "resourceName":
"intel_sriov_dpdk",
            "selectors": {
              "vendors": ["8086"],
              "devices": ["154c"],
              "drivers": ["vfio-pci"],
              "pfNames": ["enp12s0f1"]
            }
}
```



Networking Model



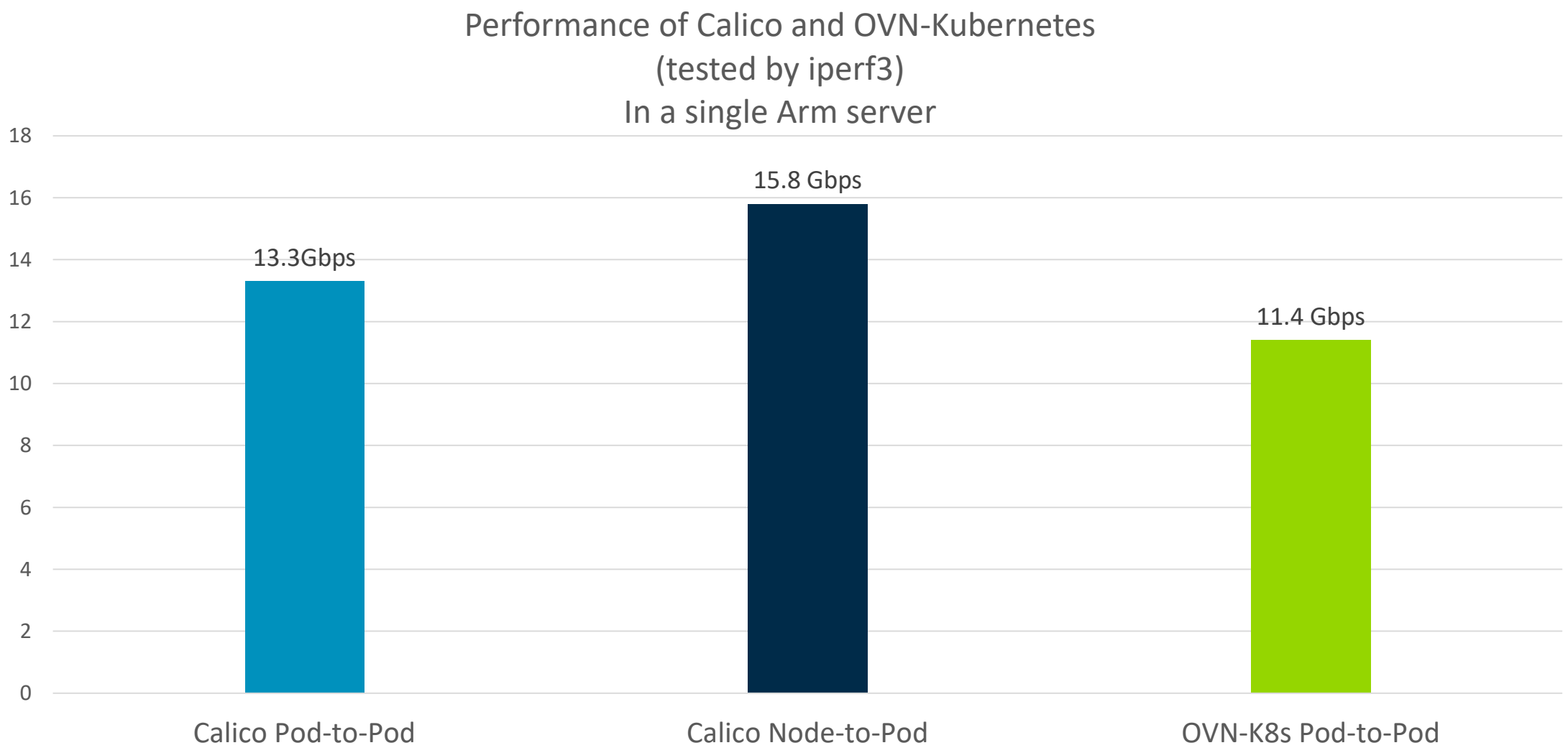
A Very Raw Thinking on the Possible Model for Container Networking with SmartNIC OVS offload support



Performance Tests

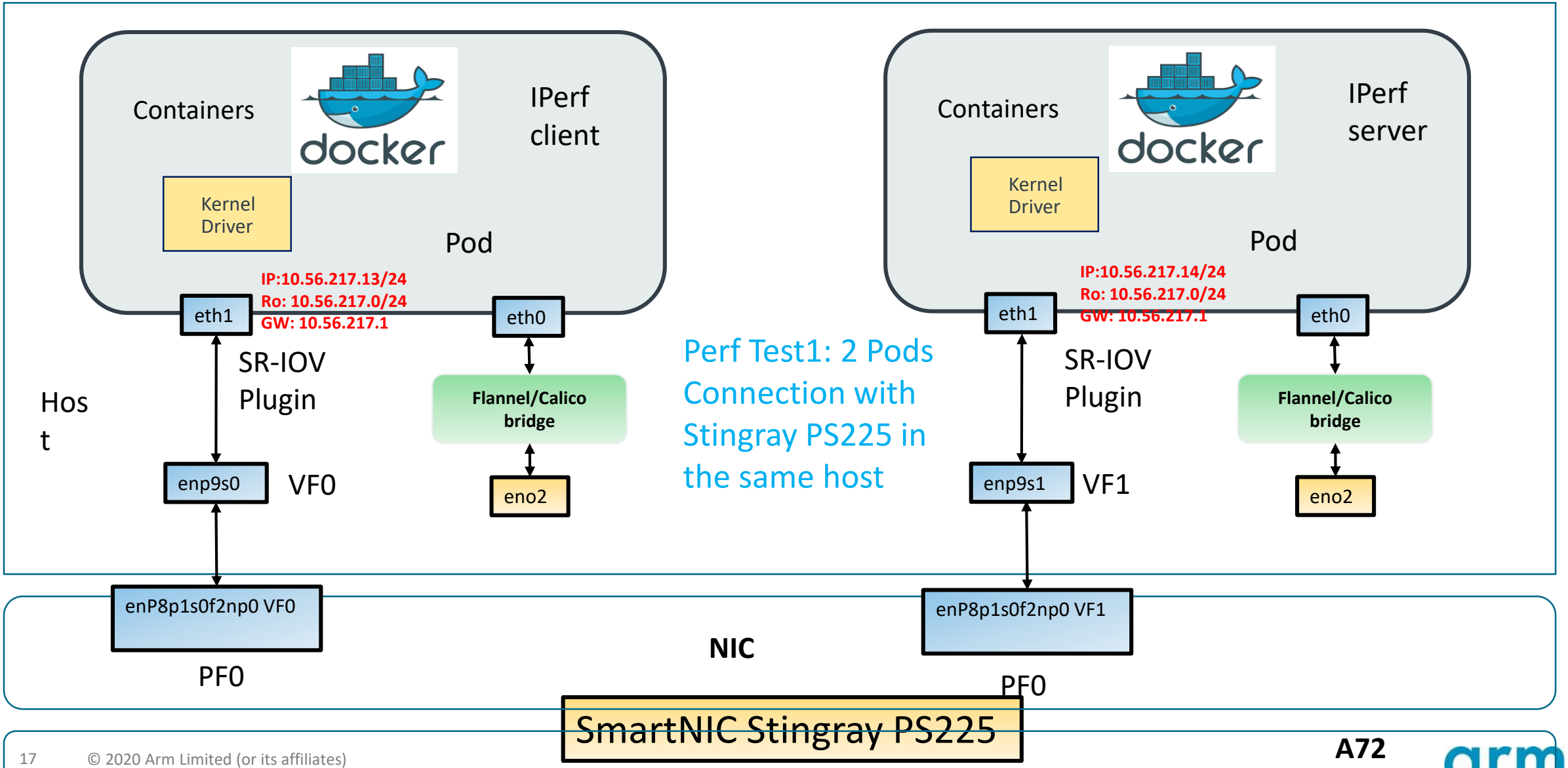
arm

Performance of Calico and OVN-Kubernetes



Performance Test with SRIOV CNI

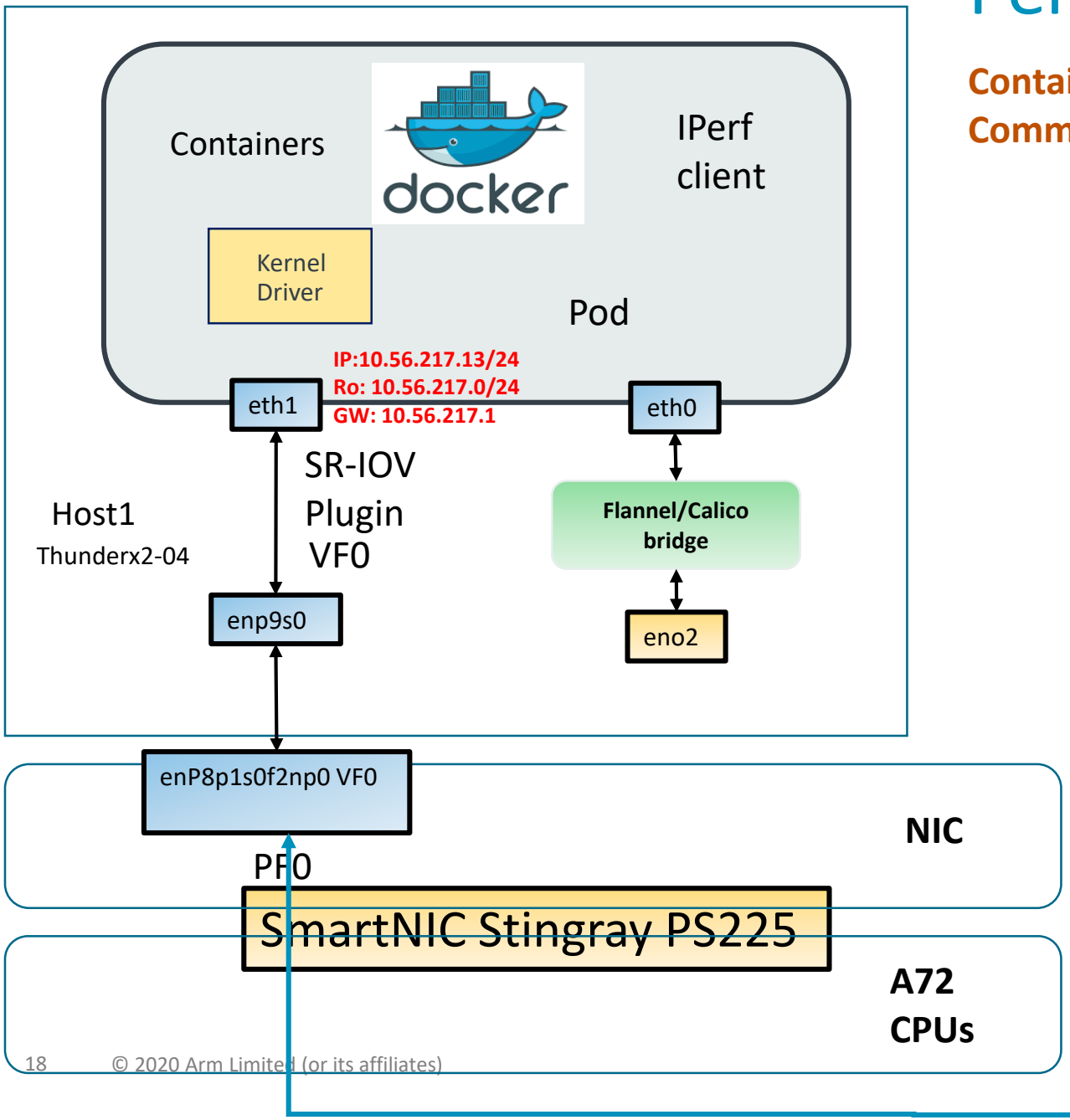
Container Networking by SRIOV Interface for Intra Host Communication



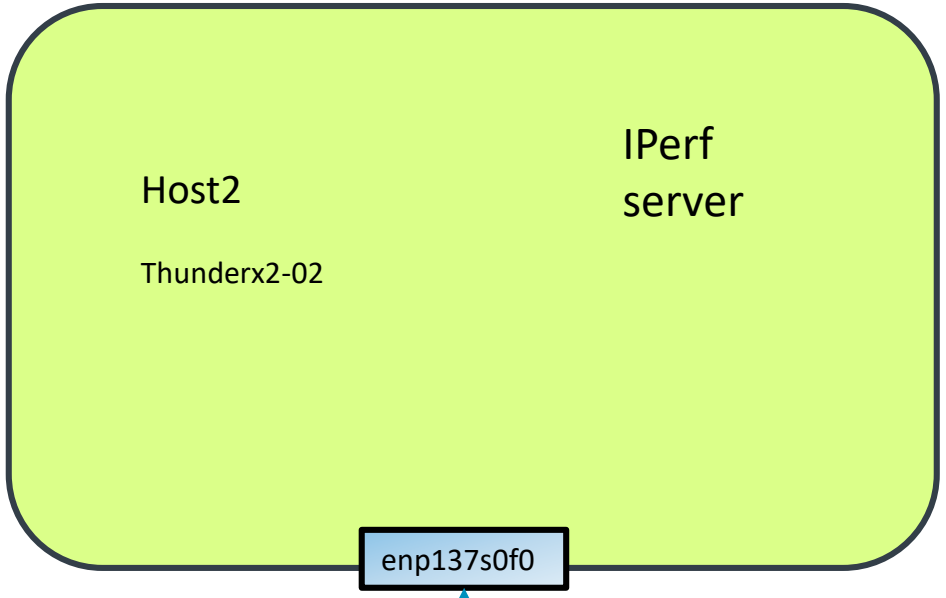
A72 CPUs

Performance Test with SRIOV CNI

Container Networking by SRIOV Interface for Inter Host Communication



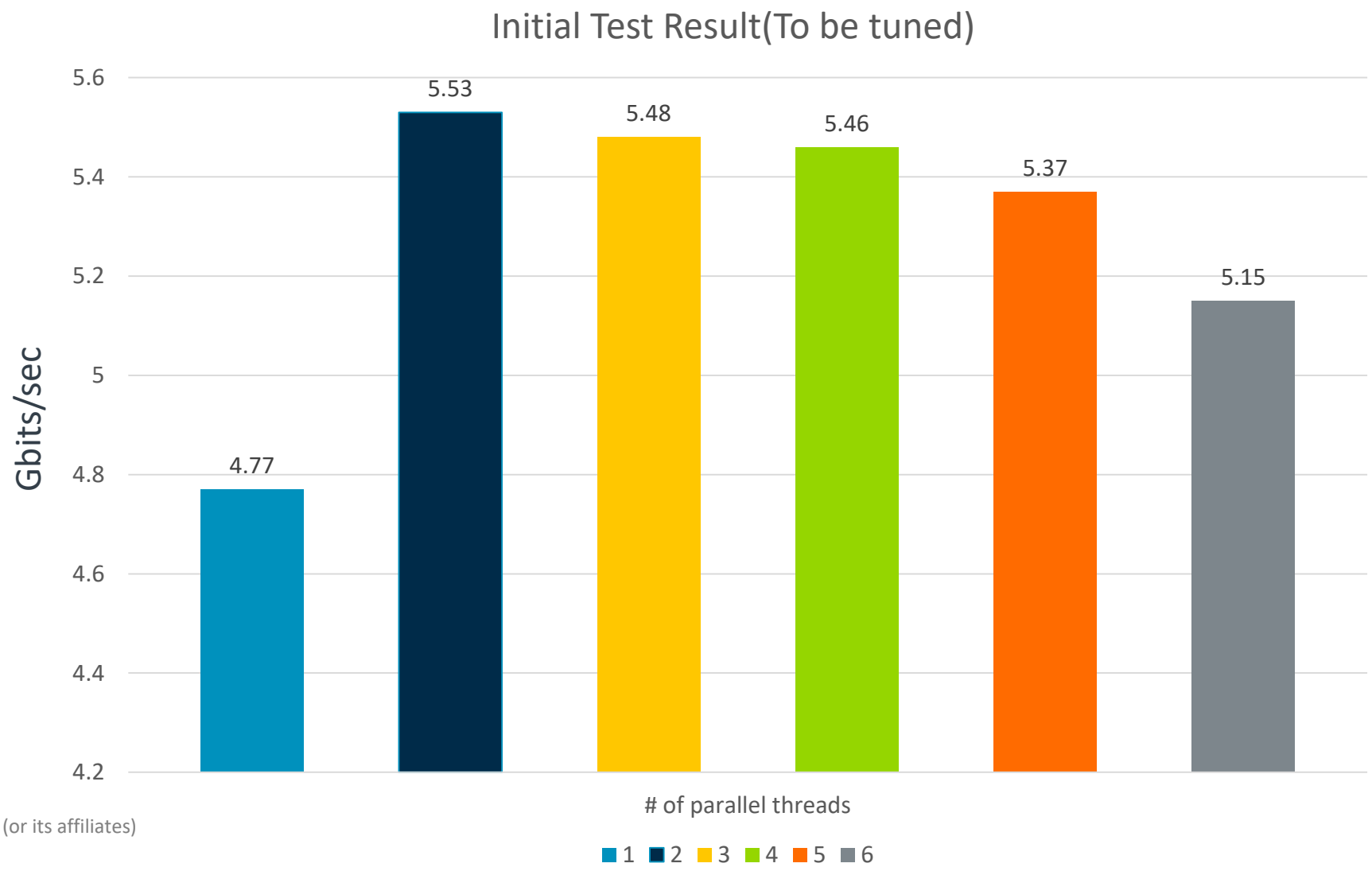
Perf Test2: 1 Pod Connected with another host via SmartNIC VF



Performance Test Result

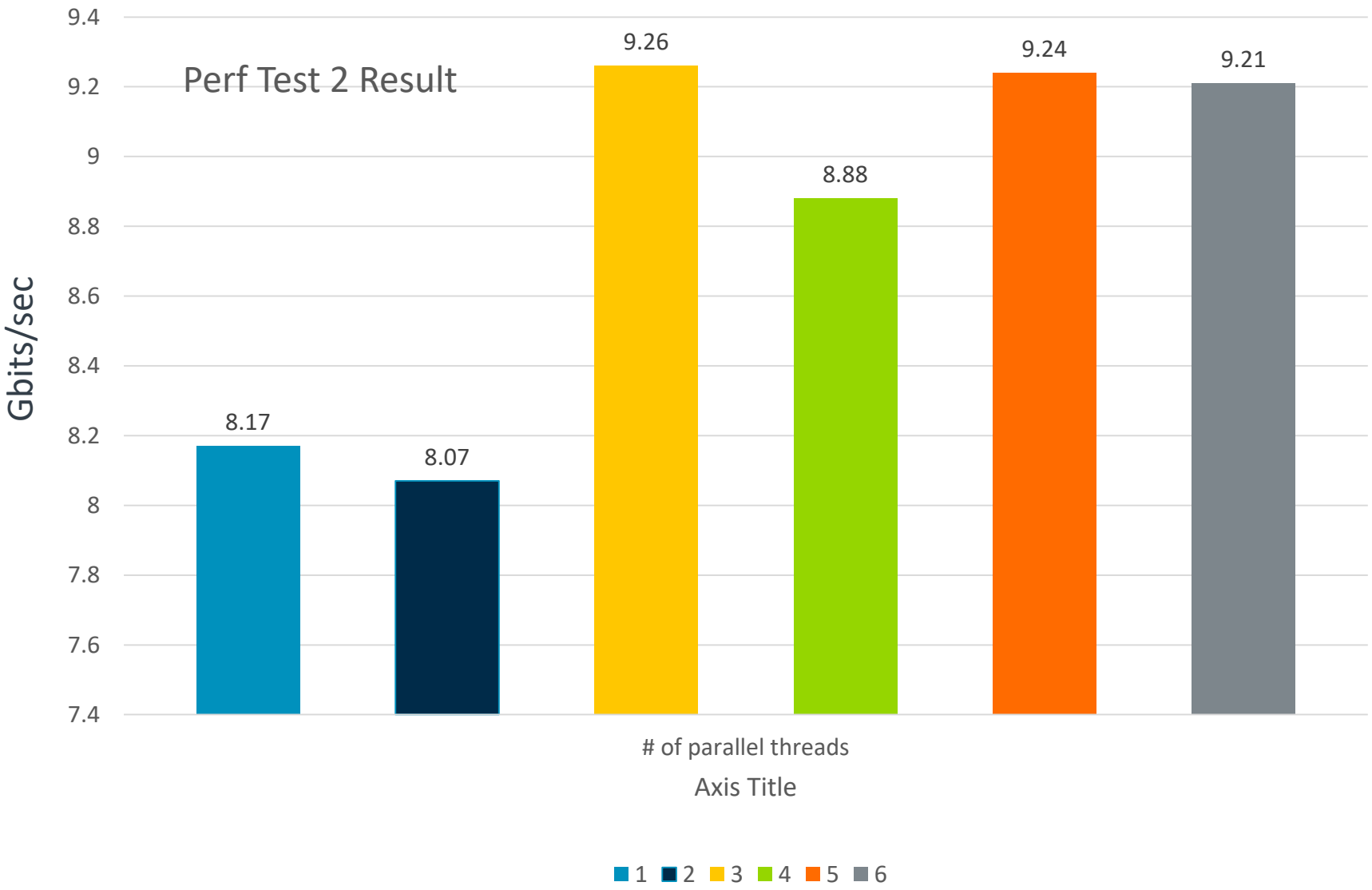
Note: Here we choose iperf instead of iperf3 due to its multi-threaded implementation

Pod Communication Performance with SRIOV CNI in the same host via PS225 VFs



Performance Test Result

Pod Communication Performance with SRIOV CNI in 2 hosts via PS225 VF (10G Connection)



Future Work(Provisional)

arm

Future Work(Provisional)

- Performance evaluation and tuning for Calico, Cilium CNI with Pod-2-Pod, Node-2-Pod cases on arm platform, mainly for cross-node. The possible tuning parameters include:
  - RPS/RFS
  - Linux system netdev parameters
  - RX/TX queue
  - MTU
  - eBPF/XDP
  - Overlay/N-Overlay
- Service mesh integration with high performance CNIs, such as Cilium/Proxy
- Further DPDK incorporated container networking usage model and performance evaluation
- Further CI for high performance container networking solutions on arm with up-to-date processes
- More thinking and possible actual work on SmartNIC support for container networking with OVS offload

arm

Thank You
Danke
Merci
谢谢
ありがとう
Gracias
Kiitos
감사합니다
धन्यवाद
شكرًا
ধন্যবাদ
תודה