Considerations for Benchmarking Network Performance in Containerized Infrastructure

draft-dcn-bmwg-containerized-infra-08

Minh-Ngoc Tran, Jangwon Lee, Younghan Kim (Soongsil University), Kyoungjae Sun (ETRI), Huynsik Yang (KT),

Draft purpose

- Distinguish benchmarking of containerized infrastructure from previous benchmarking methodology for VM-based NFV infrastructure

- Investigate **different network models that support packet acceleration technologies** based on vSwitch location considering vSwitch is the network principle of containerized infrastructure.

- Investigate **different deployment configurations** (resource isolation, hugepages, service function chaining,…) **impact** on containerized networking

Updates Summary (from v3 to v8)

- Sections 3.1 & 4 are now new section 3
- Section 3.2 is now part of new section 4
- Section 3.3 is now part of section 5
- New Appendix Sections shows our own benchmarking experiences through multiple Hackathons.

v3

v8

Detailed Updates (1)

New Section 3

Containerized Infrastructure Overview

- Comparison with VM-based Infrastructure (old 3.1)
  - The lack of hypervisior
- **Classifies different containerized deployment methods (new)**

➢ Based on that, 4 Benchmarking scenarios for the Containerized Infrastructure (old 4)

- Container2Container
- BMP2BMP
- BMP2VMP
- VMP2VMP

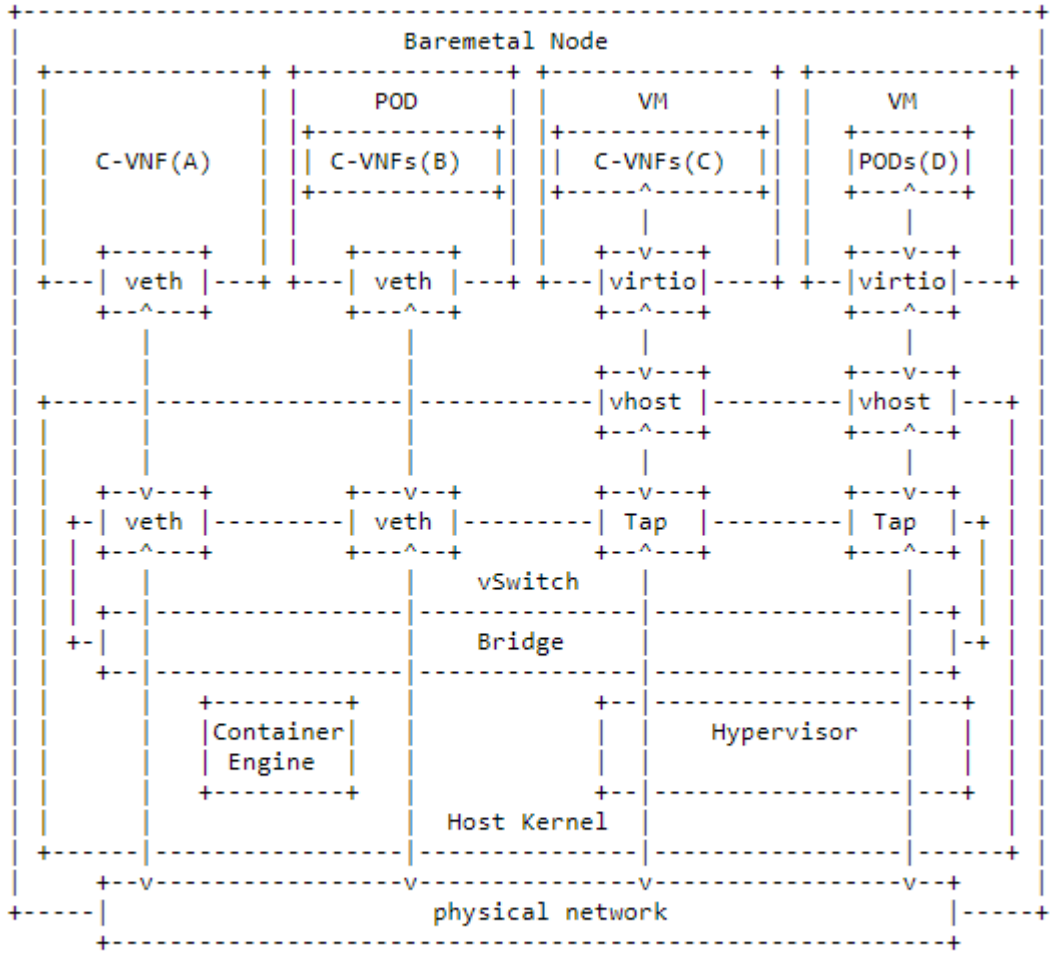


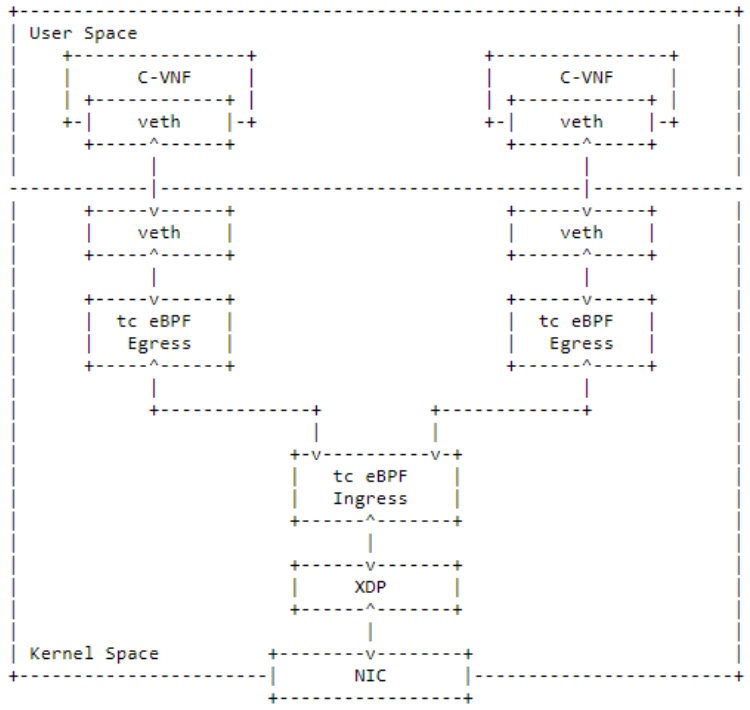
Figure 1: Examples of Networking Architecture based on Deployment Models - (A)C-VNF on Baremetal (B)Pod on Baremetal(BMP) (C)C-VNF on VM (D)Pod on VM(VMP)

Detailed Updates (2)

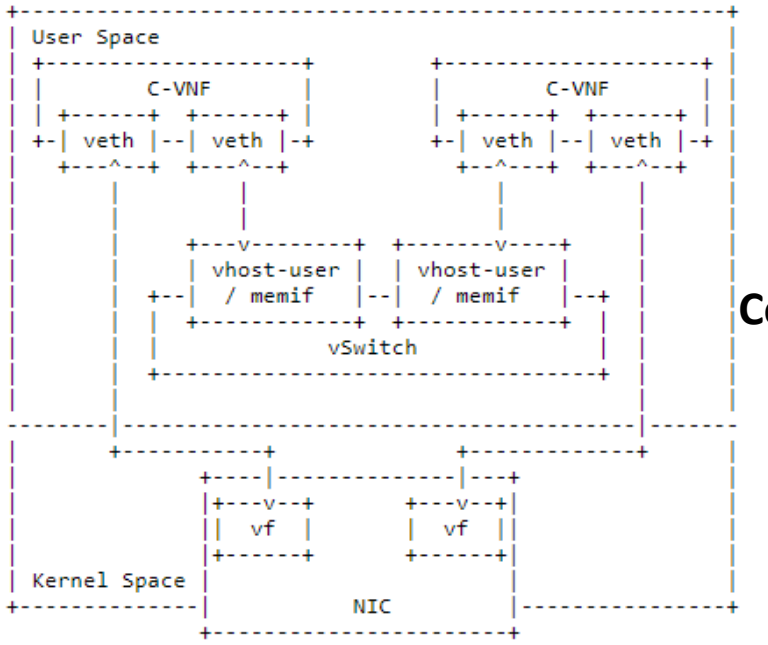
New Section 4

Networking Models in Containerized Infrastructure

- Container Networking Classification (old 3.2)
  - Kernel-space vSwitch model
  - User-space vSwitch model
  - SR-IOV (Rename to Smart-NIC Acceleration Model)
- **Adding 2 new categories**
  - eBPF Acceleration Model
    - Notable used in Cillium, Calico CNI Plugin
  - Model Combination
    - Notable used in Service Function Chaining
    - User-space vSwitch for East-West traffic
    - SR-IOV for North-South traffic



eBPF



Combination

Detailed Updates (3)

New Section 5

Performance Impacts

- Different resource considerations (old 3.3)
  - Hugepages
  - NUMA & CPU Isolation
- **Adding 2 new impacts**
  - Service Function Chaining (new 5.3)
    - In NFV environment, physical network port is commonly connected to multiple VNFs rather than single VNF
    - Aspects needed to be considered when benchmarking service function changing
      - Number of VNFs
      - Different network acceleration technologies (which provide VNF to VNF networking)
  - Inter-node networking (as new additional consideration 5.4)
    - As defined in ETSI-NFV-IFA-038, different inter-node networking technologies may affect container network performance between nodes
      - Tunnel end point (VXLAN), Border Gateway Protocol (BGP), Layer 2 underlay, direct using dedicated NIC, load balancer.

Detailed Updates (4)

New Appendix Section

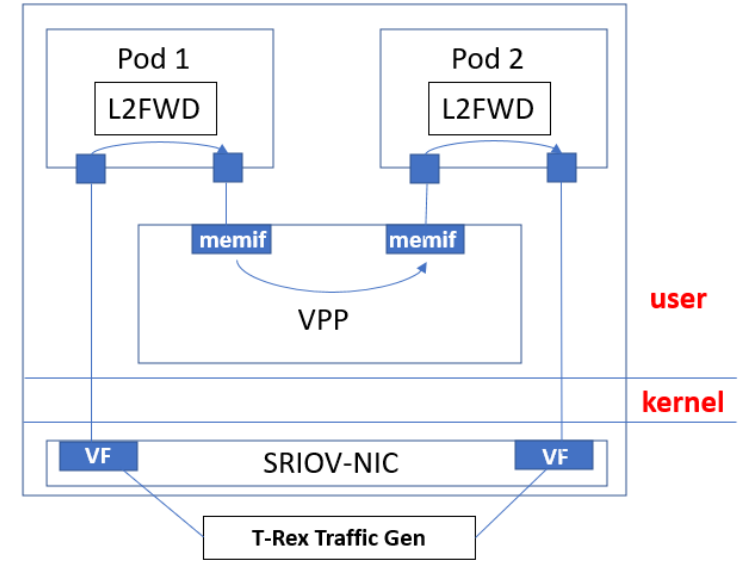
Our Hackathon benchmarking Collection

- **Different network models benchmarking**
  - VPP
  - SR-IOV
- **Different performance impacts benchmarking**
  - NUMA & CPU Isolation (included in Appendix A,B,C)
  - Service function chaining (Appendix C)

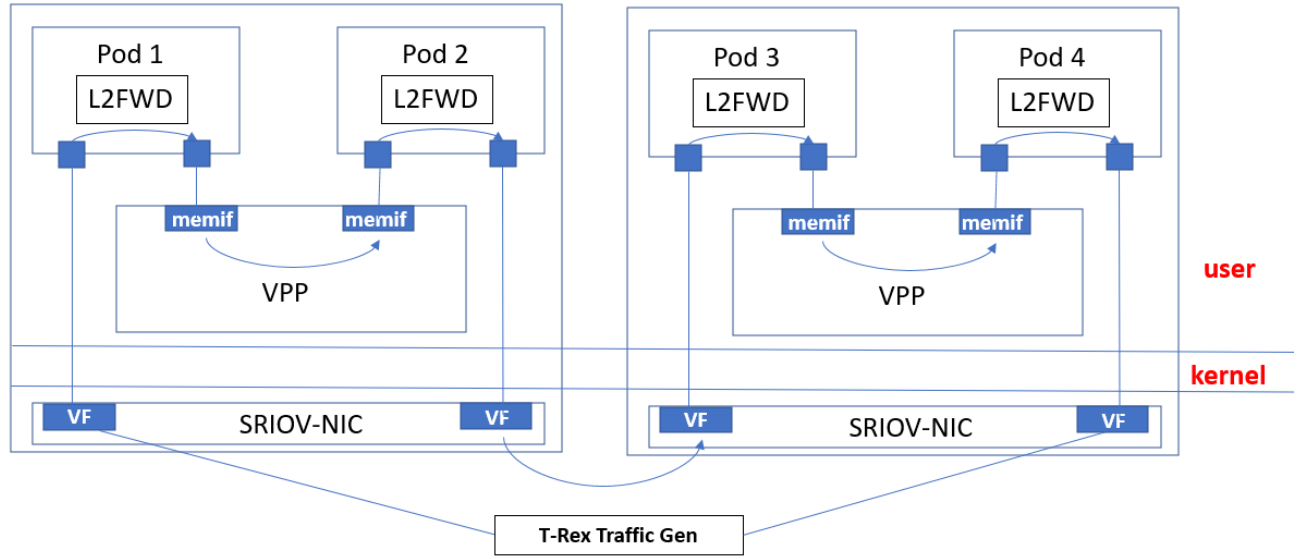
From Hackathon 113

- Service Function Chaining Benchmarking
  - Measure throughput when using SR-IOV and VPP combination

- Scenarios
  - Service Function chaining in single node
  - Service Function changing in multi-nodes (using L2 underlay as inter-node networking technique)
  - Test number of VNFs impacts (2,4,6 pods)



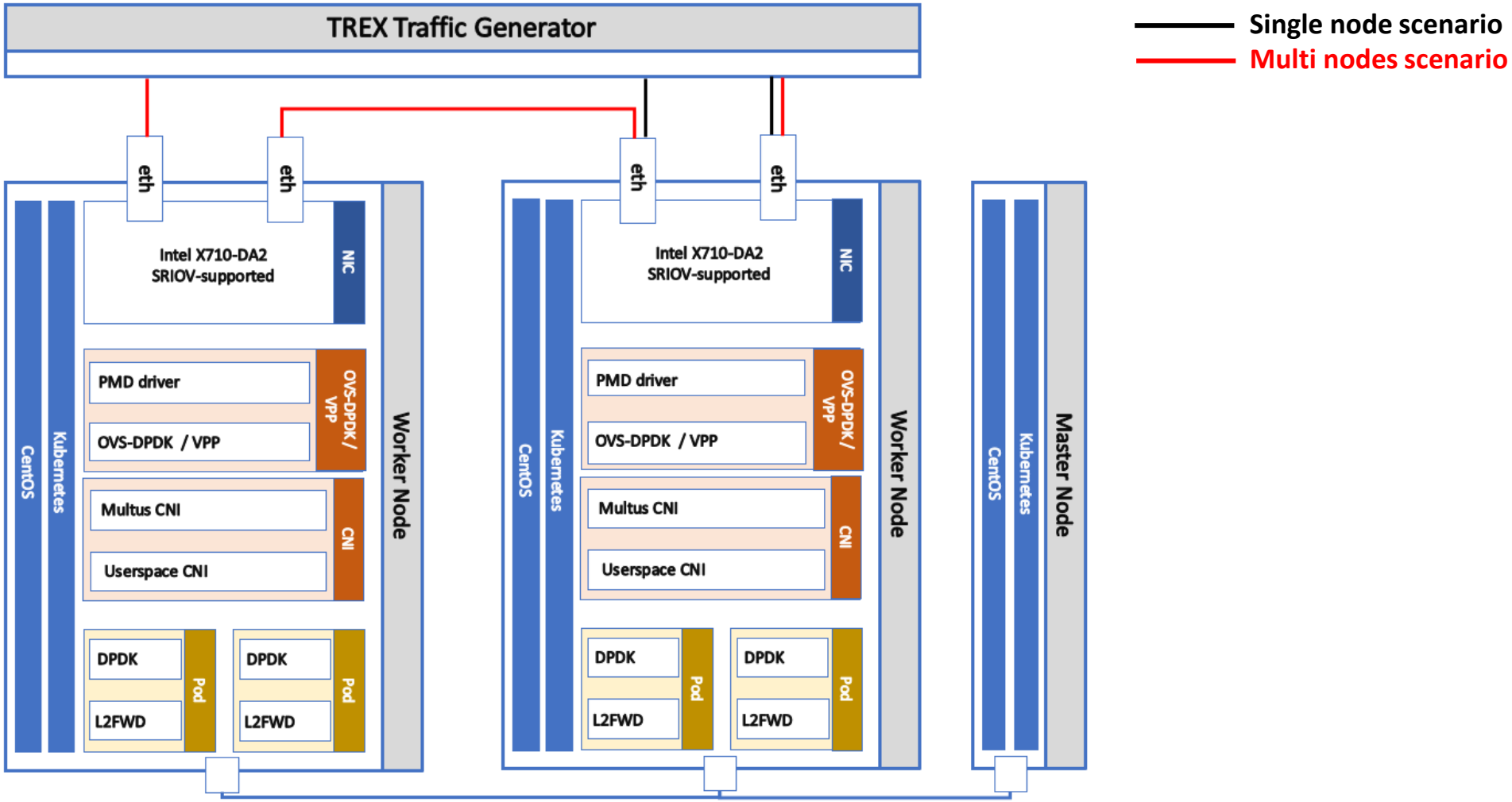
Single node scenario



Multi nodes scenario

From Hackathon 113

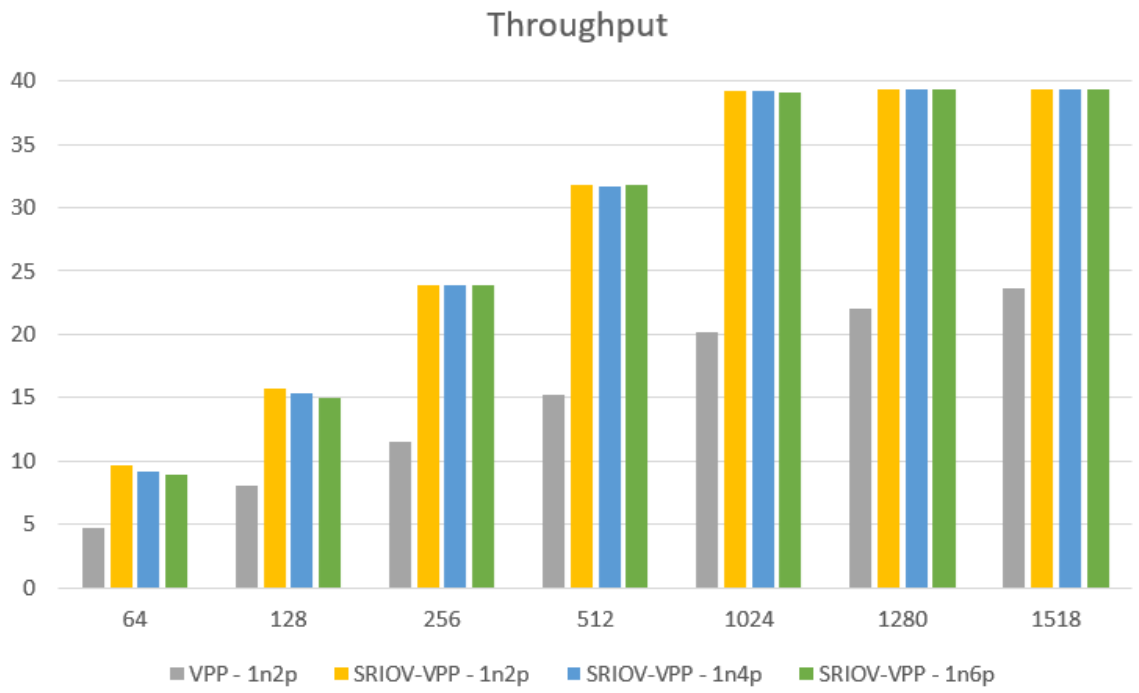
- Testbed

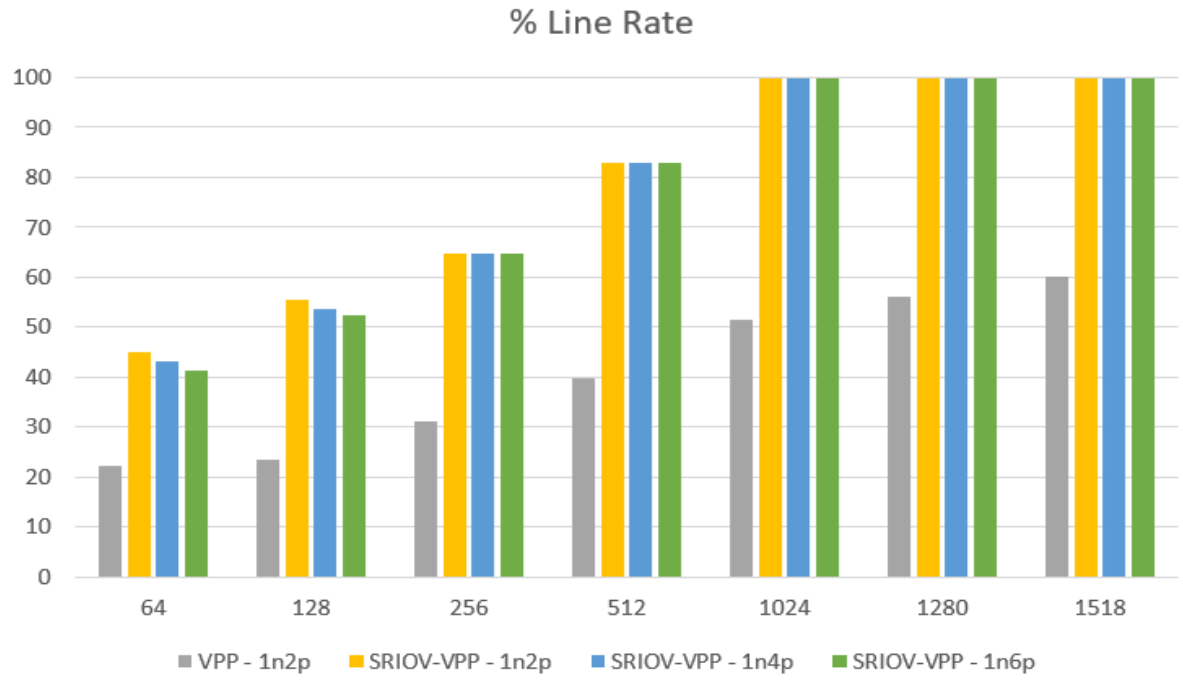


From Hackathon 113

- Benchmarking Performance Results – Single node SRIOV-VPP service chain

1. SRIOV-VPP performed significantly better than VPP only (packets through VPP need go through vSwitch, no need with SRIOV)
2. Increase number of pod slightly reduce throughput 2% at small packet size (64,128)

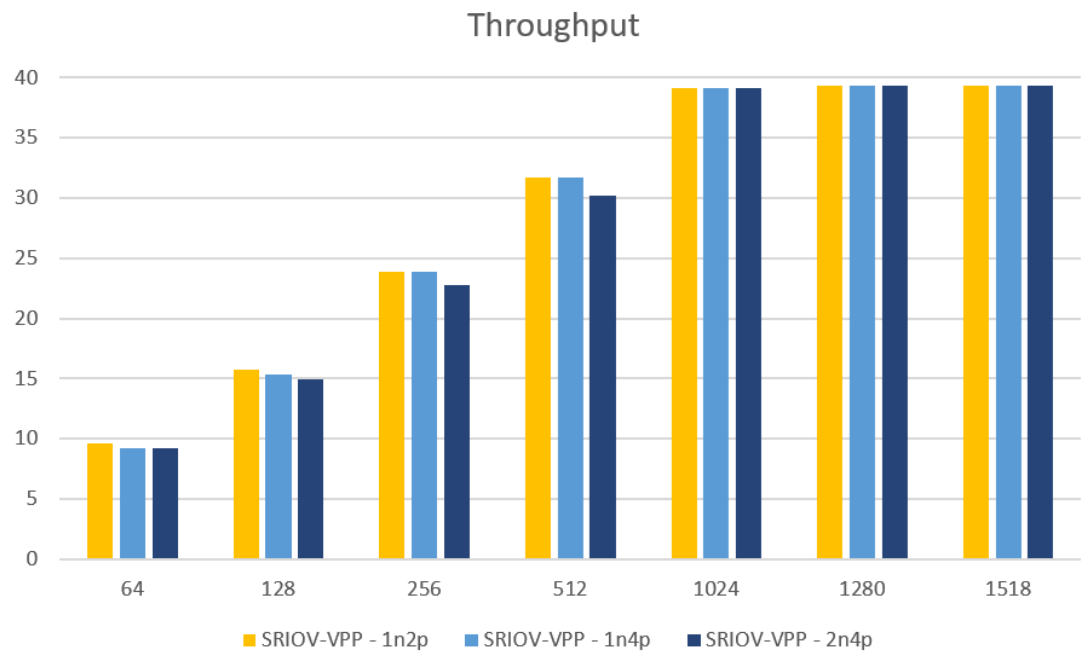


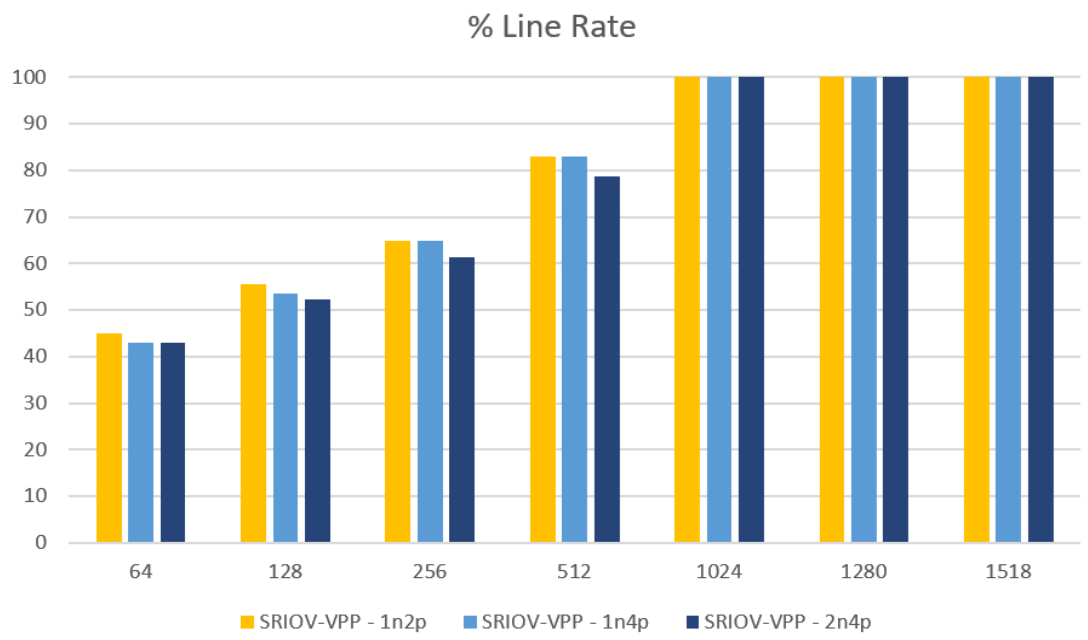


From Hackathon 113

- Benchmarking Performance Results – Multi-nodes SRIOV-VPP service chain

1. Throughput in multi-nodes scenario in slightly smaller than single-node with smaller packet size (<512) due to increasing in number of pod (4 pods total in multi-nodes > 2 pods in single node)





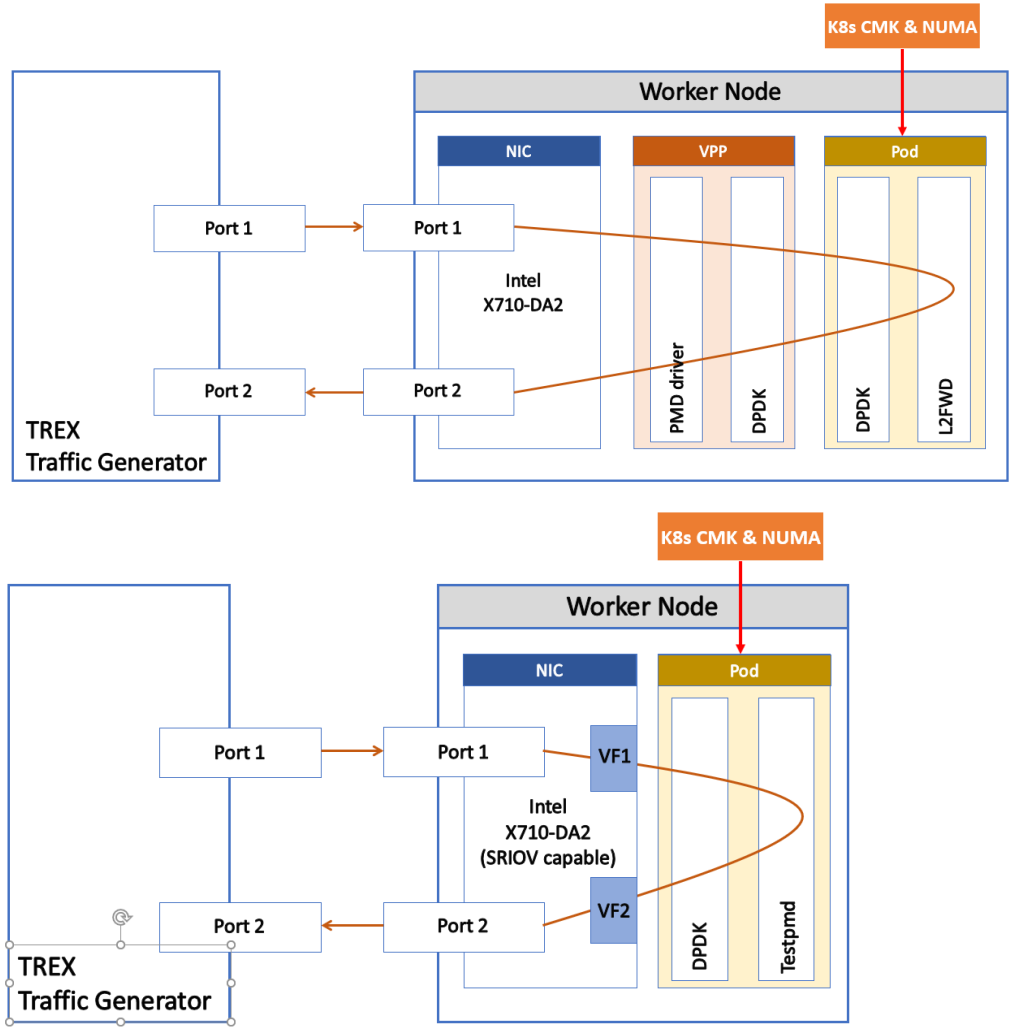
Next Steps

- Keep updating the drafts based on latest technologies
- Any comments or feedbacks are welcome
- IETF BMWG Hackathon
  - Test inter-node networking technique impacts on container network performance
  - Test performance of eBPF acceleration model (Cilium/Calico) with/without NIC offloading
  - Proof our draft scenarios and features
  - Sharing results to the BMWG

Backup Slides

Benchmarking Experiences (Contiv-VPP + SRIOV)

- Test performance of user-space based model and SmartNIC (VPP and SRIOV)
- Figure out impact of CPU isolation (using CMK – CPU Manager for Kubernetes) and NUMA to network performance
  - Without CMK
  - CMK-shared mode (2 pods share 2 CPUs)
  - CMK-exclusive mode (1 dedicated CPU/pod)



Benchmarking Experiences (Contiv-VPP + SRIOV)

What we learned

- VPP and SRIOV has nearly the same performance

CPU Isolation:

- CPU Isolation (CMK) significantly improves throughput
- Exclusive mode is better than Shared mode

NUMA alignment:

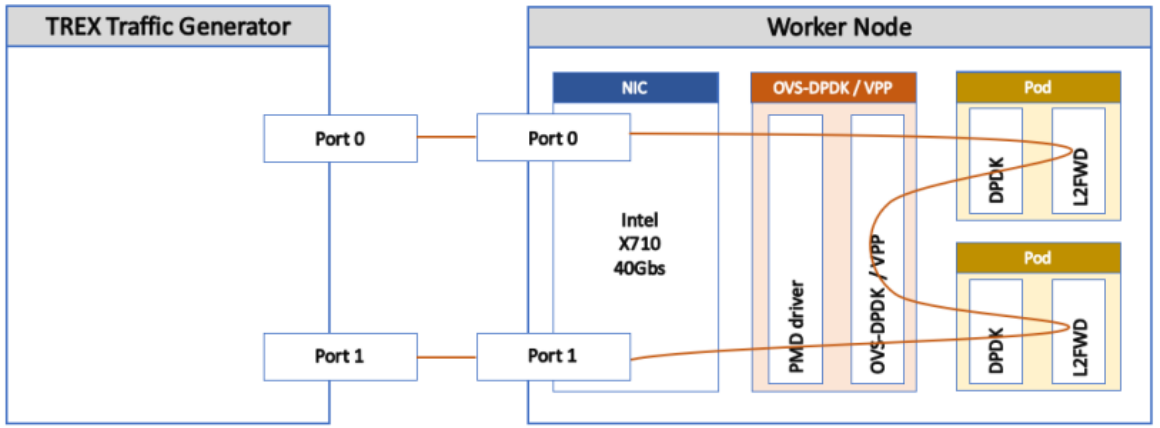
- Assigning CPU in the same NUMA node is better than in different NUMA nodes

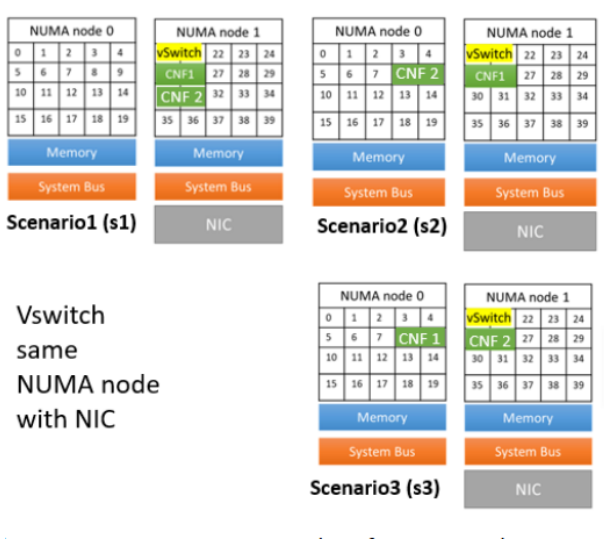
Model	NUMA Mode (pinning)	Result(Gbps)
Maximum Line Rate	N/A	3.1
	same NUMA	9.8
Without CMK	N/A	1.5
CMK-Exclusive Mode	same NUMA	4.7
	Different NUMA	3.1
CMK-shared Mode	same NUMA	3.5
	Different NUMA	2.3

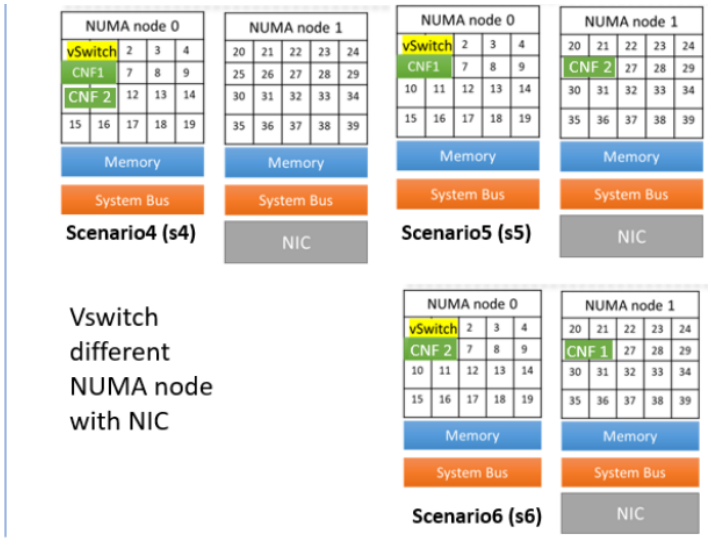
CPU Isolation and NUMA location impact in VPP test with 10G Intel X710-DA2 NIC

Benchmarking Experiences (Multi-pods)

- Test performance of VPP in service function chain scenario (2 pods)
- Figure out impact of NUMA allocation over CNF, vSwitch, NIC
  - 6 scenarios
    - vSwitch same with NIC
      - vSwitch same with input CNF and vice versa
    - vSwitch different with NIC
      - vSwitch same with input CNF and vice versa







Benchmarking Experiences (Multi-pods)

What we learned

NUMA alignment:

- **vSwitch and NIC** in different nodes slightly degrade performance in 1024+ packet size
- **CNFs and vSwitch** in different nodes degrade performance by 10-15%
- **Input CNF and vSwitch** in different node has better performance