LF NETWORKING

LFN Developer & Testing Forum

K8S Container Networking Benchmarking

Sridhar (Team-VSPERF)

sridhar.rao@spirent.com

@ngignir

Thanks…

- We were/are not experts in K8S or K8S Networking.
- It has taken lot of inputs and suggestions from different community members.
- Just to name a few … Thanks a lot!
  - Kuralamudhan Ramakrishnan @ Intel
  - Billy McFall @ Redhat
  - Douglas Smith @ Redhat
  - Martin Klozik @ TietoEvry

The flow

- Share experience (and results) based on the ongoing work.
  - 8 Months+
- Summarize the results in points rather than walking through all the plots.
  - 20+
- Open for feedback and inputs.

VSPERF

- Tooling and Testcases for Dataplane Performance Testing and Benchmarking.
- Variations: Baremetal, Openstack and Kubernetes.
- Low-Granular Configuration and Customizations, Commercial Traffic Generators with opensource Interface on Open Testbeds
- "Experimental in Nature" -- Efforts will go as inputs to Standardization (de jure or de facto) or Publications.

VSPERF

Pre-Deployment

Post-Deployment

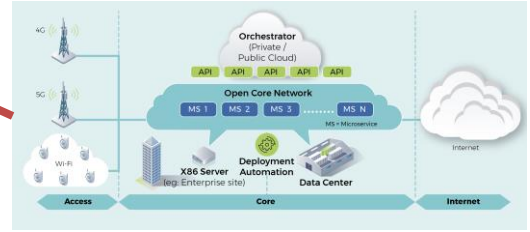
Anuket

VSperf

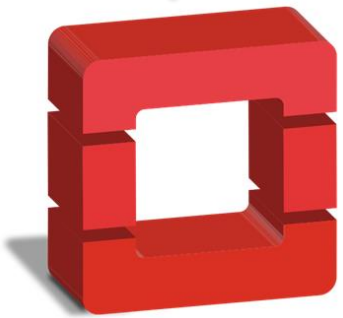
Dataplane performance testing and benchmarking

Test-Lab

Production Network



Baremetal



Openstack

Kubernetes

The Name of the Project, considering its current scope, is indeed misleading. This project will be renamed soon.

Framework Capability & Experimental Studies

LFN Developer & Testing Forum

VSPERF Framework Capability

VSPERF Experimental Studies.

L2 Cache Management with Intel RDT ← Use of L2Cache management on VNF Performance

RFC2544 With Loss-Verification ← RFC2544 Binary-Search algorithm for virtualized environments

Numa Management, “live-Results”, Metrics Correlation, etc. ← Cross-Numa Performance Studies

Automated K8S Cluster, DUT & Test management. ← K8S CNB for Telco Usecases

K8S CNB: Framework Capability

- Automated Cluster Setup
- Automated Test Setup
  - DUT, TGen, Metrics, etc.
- Automated Test Runs *

Automated Cluster Deployment

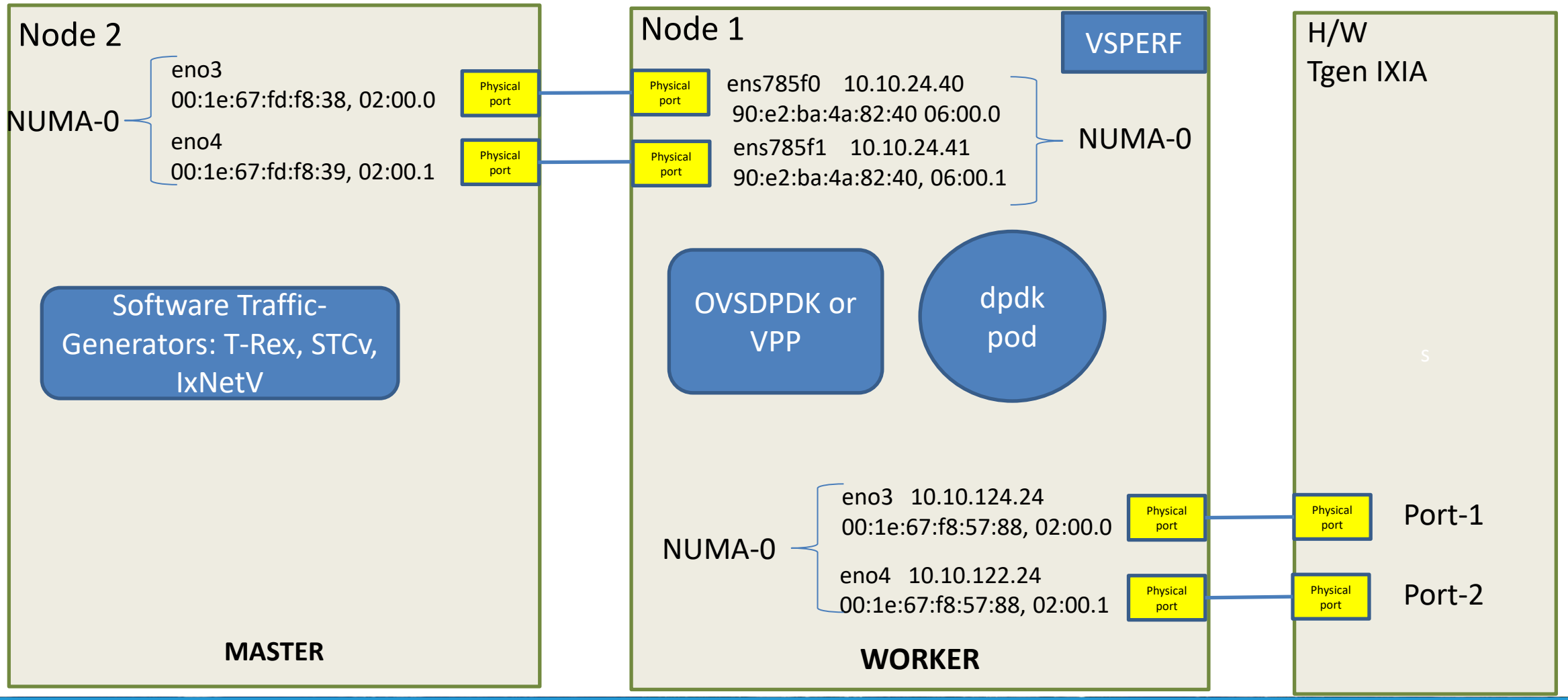
- Ansible Role.
- Hosts:
  - Master(s)
  - Worker(s)
- CNIs
  - Multus, flannel, Userspace, SRIOV, etc..
- Device Plugin
  - SRIOV

Automated Test Setup

- Vswitch – in case of OVSDPDK or VPP.
- VFs – In case of SRIOV.
- Network Attachment
- Pod Definition.
- Enable Forwarding in Pod.

K8S CNB : Experimental Study

LFN Developer & Testing Forum



K8S CNB : Experimental Study

LFN Developer & Testing Forum

Feature	Values
Kubernetes Version	1.18
Docker Version	19.03.12
CNIs	Multus, Flannel, Userspace
Device Plugin	SRIOV
Default CNI	Flannel
OVSDPDK: Vswitch CPU	2
OVSDPDK PMD CPU	3, 4, 5, and 6
DPDK Version	19.02 and 20.03
Forwarding Application	L2FWD and TESTPMD
CPUs for POD	7,8,9 & 10
VFs per NIC	1
PMD Driver	VFIO_PCI

OVS-Version	2.12.0
VPP-Version	19.04.4
TREX Version	2.48
VSPERF Release	Iruya
NIC Speed	10Gig
NICS	Intel 82599
Userspace CNI Version	1.2
SRIOV Network Device Plugin version	3.2
Server Capacity	44 CPUs. SRIOV-Capable NICS. 138GB RAM. 1TB HD.
CPUs allotted to Traffic Generators	6
Default memory and CPU for the POD	4GB and 4CPU.

K8S CNB : Experimental Study

- What are the K8S networking Performance study Gaps?
- For DPDK-based multi-interface pods, what are the application CNIs.
- What are the challenges in Setting up CNIs?
- What are the performance differences among Userspace CNI OVSDPDK, and VPP, in comparison with SRIOV CNI?
  - Different Traffic Patterns.
- What really affects the performance?
  - Performance Tuning:
    - CPU Count, Memory Size, CPU Isolation, NUMA-awareness, Multi-Queue.

K8S CNB: Experimental Study

- K8S networking Performance study
  - Literature Studies (academic publications).
  - Performance studies by opensource projects.
  - Performance studies (independent articles).
  - Clear gap w.r.t Telco-Usecase (multi-interface, acceleration techniques, Standards-driven tests).
- SRIOV, Userspace CNI (OVSDPDK, VPP).
- DPDK Version (20.05), Userspace cni (ovsdpdk with memif), multi-queue configuration.
- VPP with memif outperforms OVSDPDK with vhostuser
  - Significant with lower-sized packets.

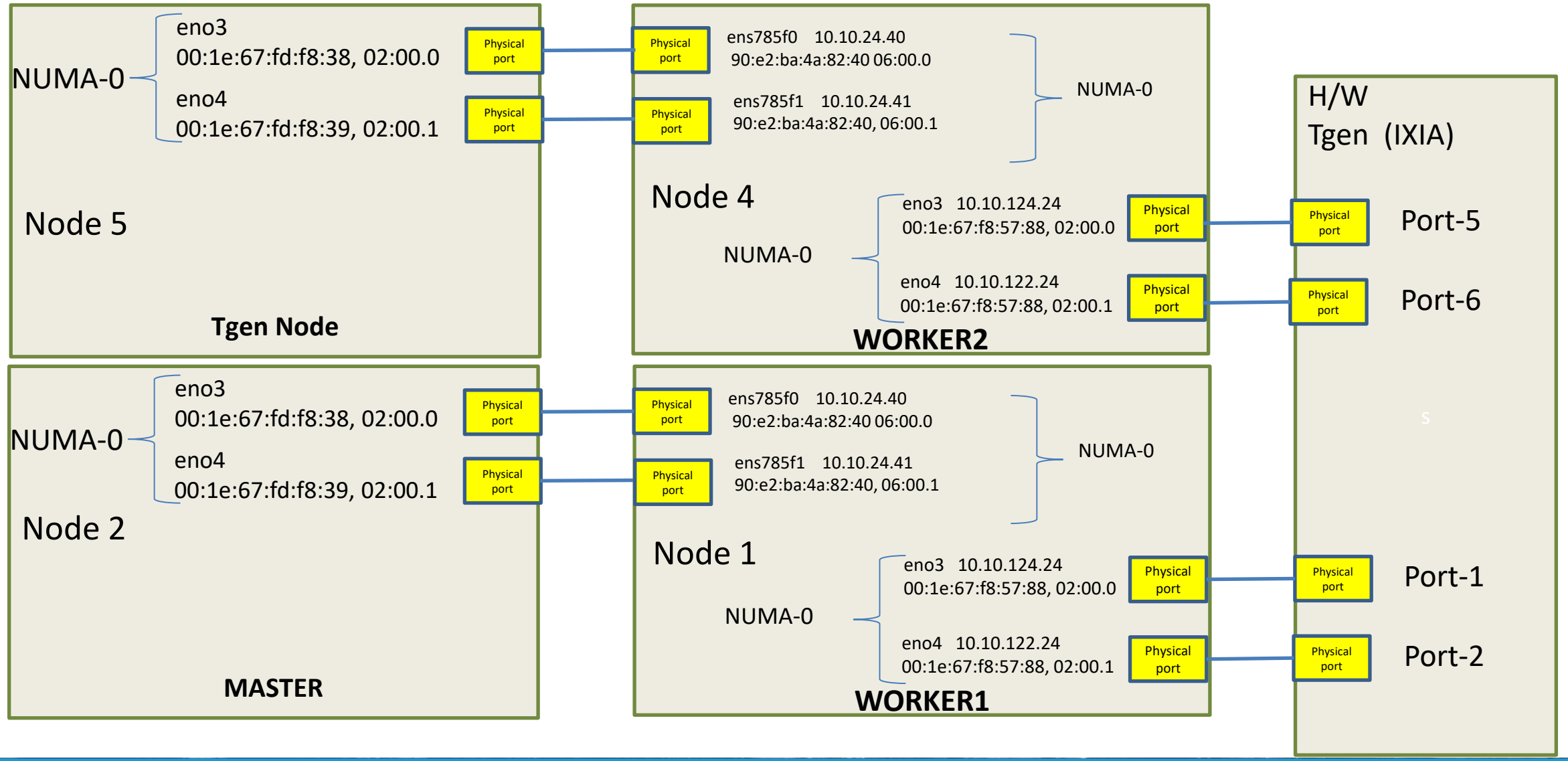
K8S CNB: Experimental Study

- Binary-search (without loss-verification) can make difference.
  - Better to ‘loss-verification’ approach.
- RFC2544 Back2Back studies provides better insights.
- In general, Memif is preferred over vhostuser
- CPU count Memory Size.
  - Only upto a certain number, beyond which it has no impact
  - For the considered pods, 6 CPU and 6GB RAM.
- Numa-alignment has significant impact.
  - vSwitch and Interfaces in Numa0 and Pods in Numa1 – resulted in performance degrade of 10 – 35%

Next Steps

- OVSDPDK with memif
- MultiPod
- Multihost, Multipod
- Tgen Pods
  - Commercial (Spirent, Keysight) and Opensource (T-Rex, DPPD-Prox)
- Other CNI
  - Open for suggestions.
- *(vSwitches in a Container, DanM vs Multus)*
  - Performance may not be affected.

Next Steps


NUMA-0
eno3
00:1e:67:fd:f8:38, 02:00.0
eno4
00:1e:67:fd:f8:39, 02:00.1
Physical port
Physical port
Node 5
Tgen Node
Physical port
Physical port
ens785f0 10.10.24.40
90:e2:ba:4a:82:40 06:00.0
ens785f1 10.10.24.41
90:e2:ba:4a:82:40, 06:00.1
NUMA-0
Node 4
NUMA-0
eno3 10.10.124.24
00:1e:67:f8:57:88, 02:00.0
eno4 10.10.122.24
00:1e:67:f8:57:88, 02:00.1
Physical port
Physical port
WORKER2
H/W
Tgen (IXIA)
Physical port
Physical port
Port-5
Port-6
NUMA-0
eno3
00:1e:67:fd:f8:38, 02:00.0
eno4
00:1e:67:fd:f8:39, 02:00.1
Physical port
Physical port
Node 2
MASTER
Physical port
Physical port
ens785f0 10.10.24.40
90:e2:ba:4a:82:40 06:00.0
ens785f1 10.10.24.41
90:e2:ba:4a:82:40, 06:00.1
NUMA-0
Node 1
NUMA-0
eno3 10.10.124.24
00:1e:67:f8:57:88, 02:00.0
eno4 10.10.122.24
00:1e:67:f8:57:88, 02:00.1
Physical port
Physical port
WORKER1
Physical port
Physical port
Port-1
Port-2

LF NETWORKING

LFN Developer & Testing Forum

- Open for Feedback/Inputs
- Join the Project
- Join VSPERF meetings (Wednesday 8AM PT).
- sridhar.rao@spirent.com