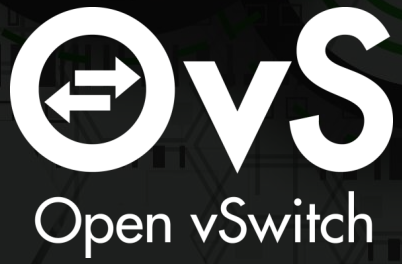


Mark Kavanagh / Tarek Radi

Intel Corporation

Optimizing TCP Workloads in an OvS-based NFV Deployment

Legal Disclaimer

General Disclaimer:

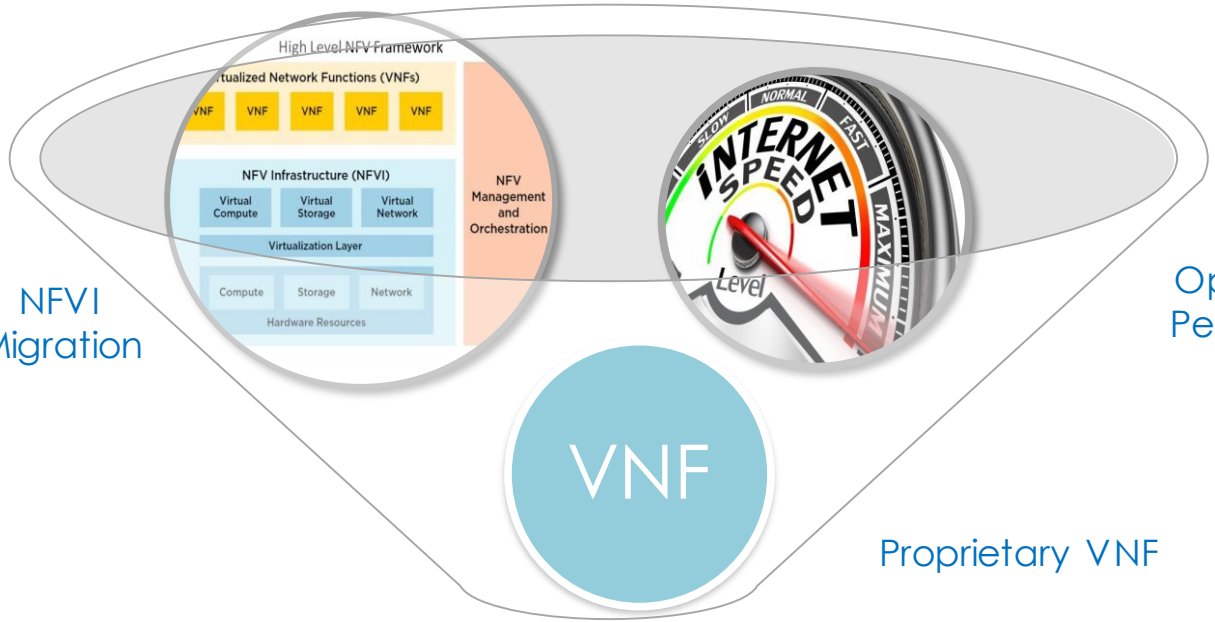
Technology Disclaimer:

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Performance varies depending on system configuration. No computer system can be absolutely secure. Check with your system manufacturer or retailer or learn more at [intel.com].

Performance Disclaimers:

Cost reduction scenarios described are intended as examples of how a given Intel-based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction. Results have been estimated or simulated using internal Intel analysis or architecture simulation or modelling, and provided to you for informational purposes. Any differences in your system hardware, software or configuration may affect your actual performance.

Problem Domain



NFVI Migration

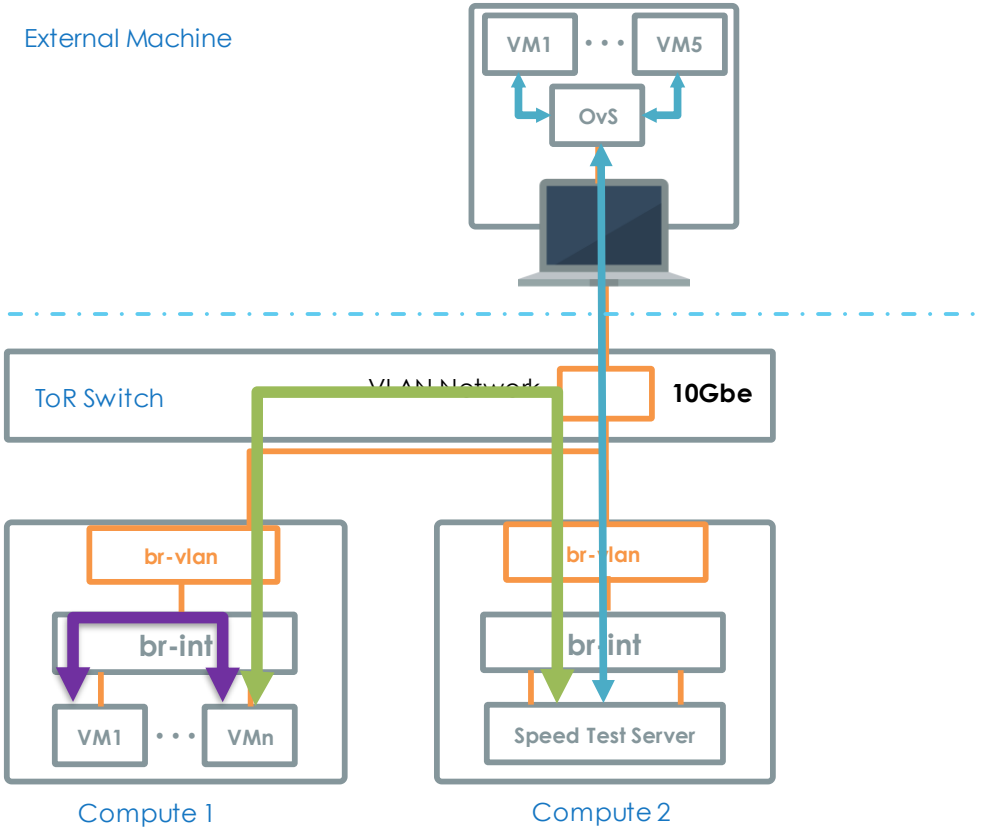
Optimal TCP Performance

Proprietary VNF

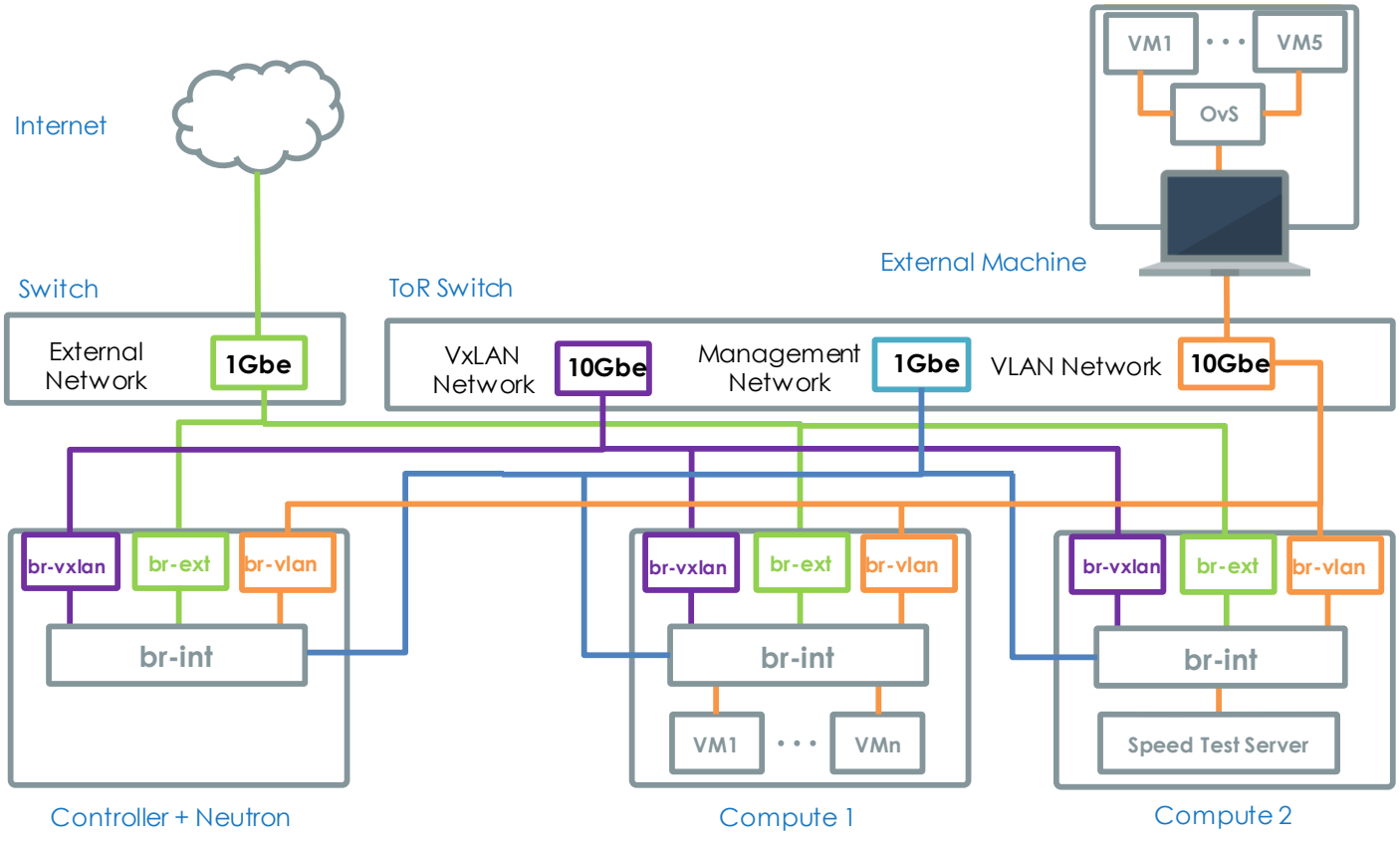
Deployment Scenario (Simplified)

Customer-Defined Test Cases

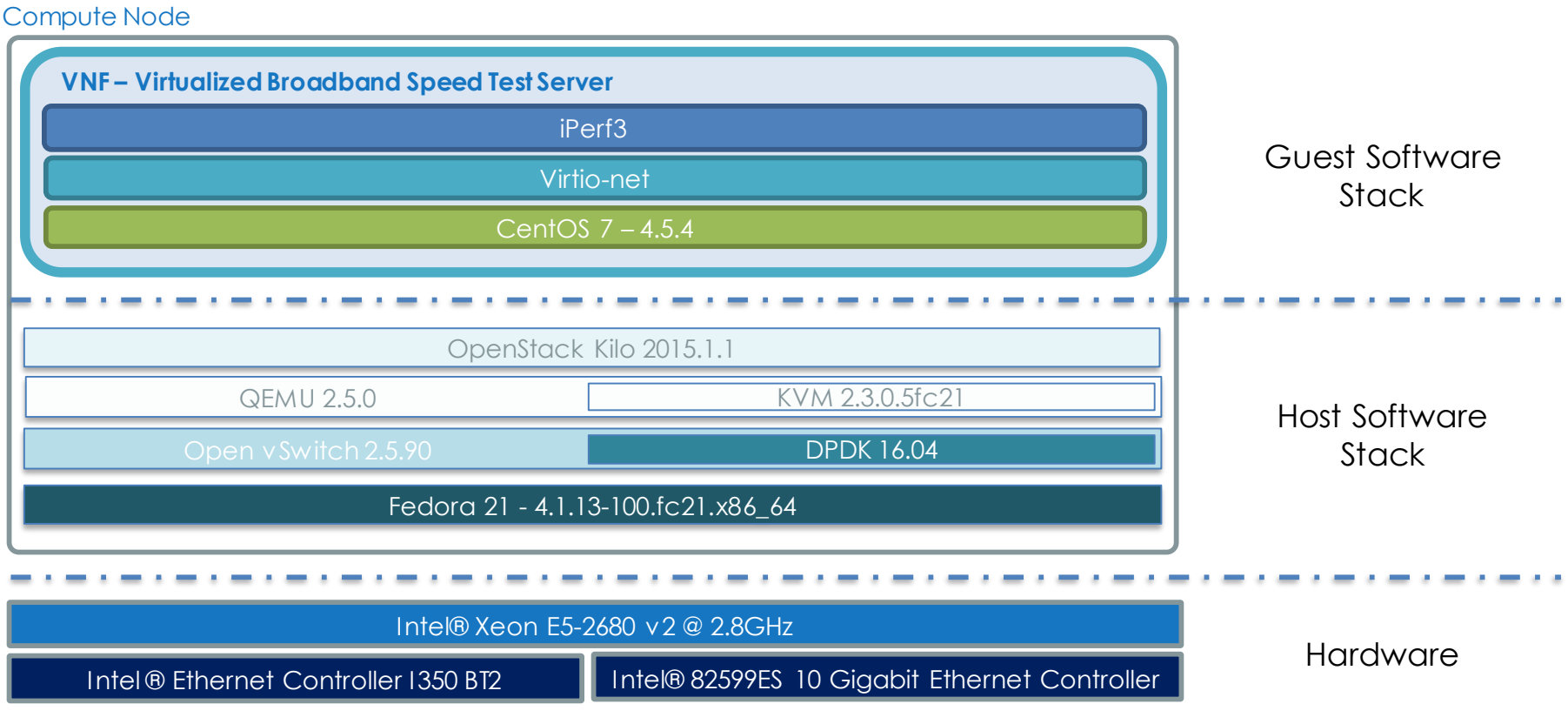
Speed Test Client	→	Speed Test Server
VM *External Network*	→	Speed Test Server *Compute Node*
Speed Test Server *Compute Node 2*	→	VM *Compute Node 1*
VM *Compute Node 1*	→	VM *Compute Node 1*



VNF Deployment Scenario (Full)



Anatomy of a VNF Compute Node



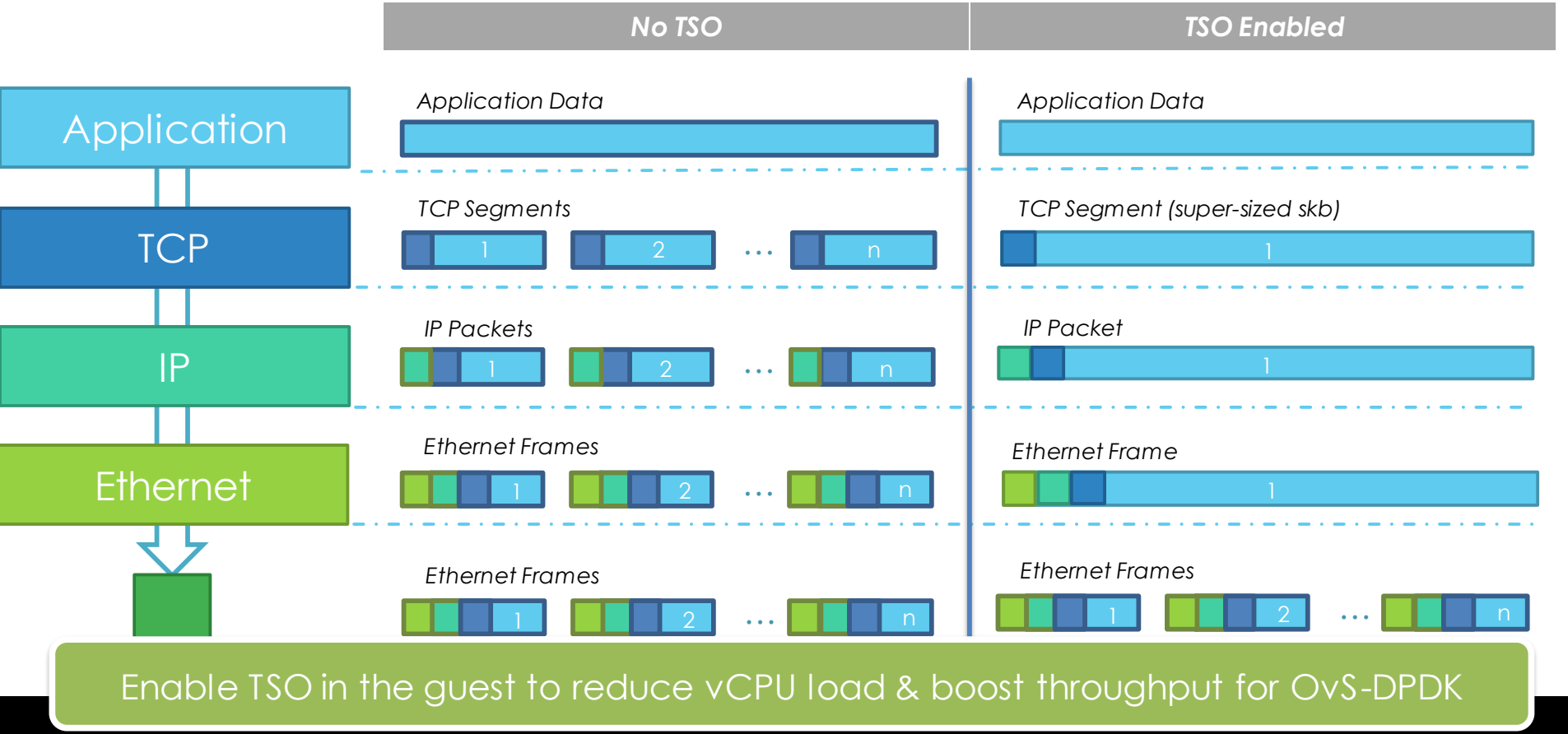
Optimizations: Baseline

- ✓ Enable Hugepages
  - *Reduce the impact of Translation Lookaside Buffer (TLB) misses*
- ✓ Affinitize DPDK PMDs, and QEMU's virtual CPU threads
  - *Maximize CPU occupancy*
  - *Minimize cache thrashing*
- ✓ Enable NUMA support for OvS-DPDK
  - *Eliminate QPI traversal performance penalties*

Additional details available here

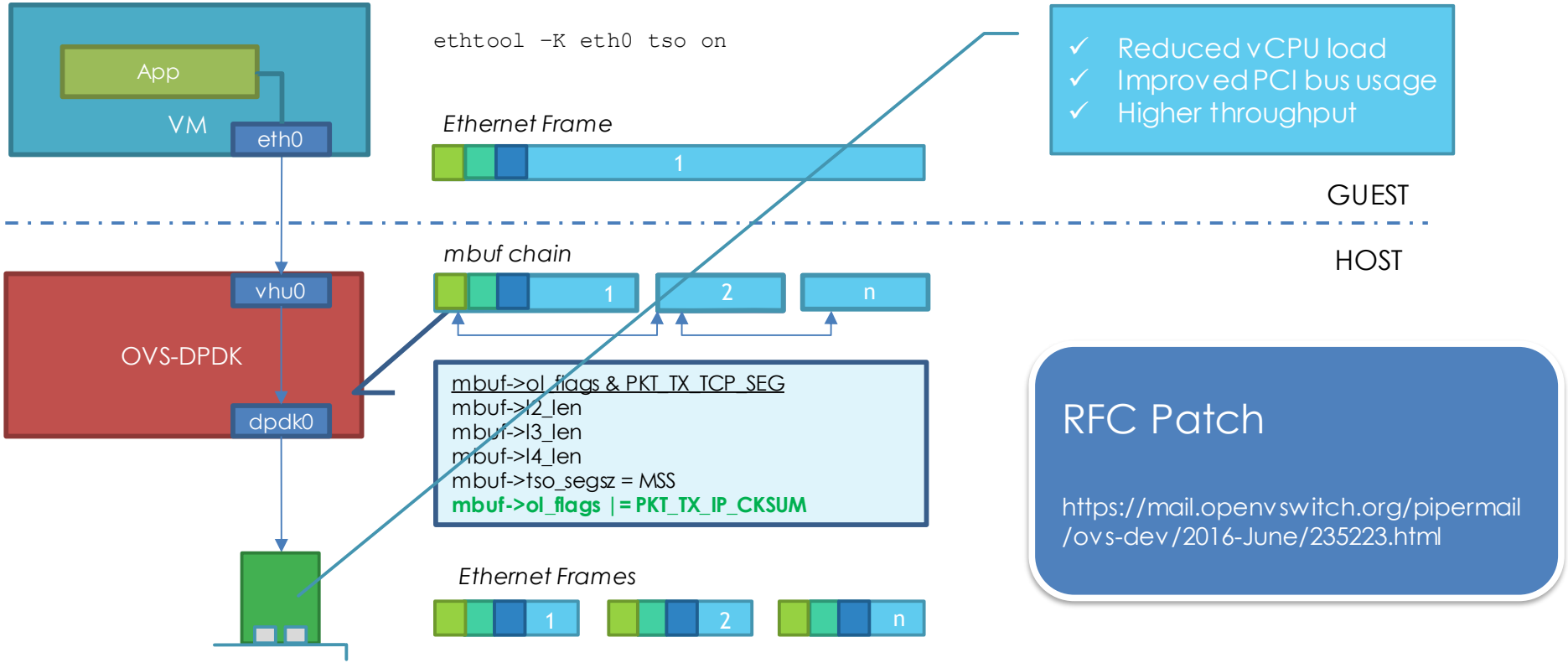
https://github.com/openvswitch/ovs/blob/master/INSTALL.DPDK-ADVANCED.md

Optimizations: TCP Segmentation Offload (TSO) Overview

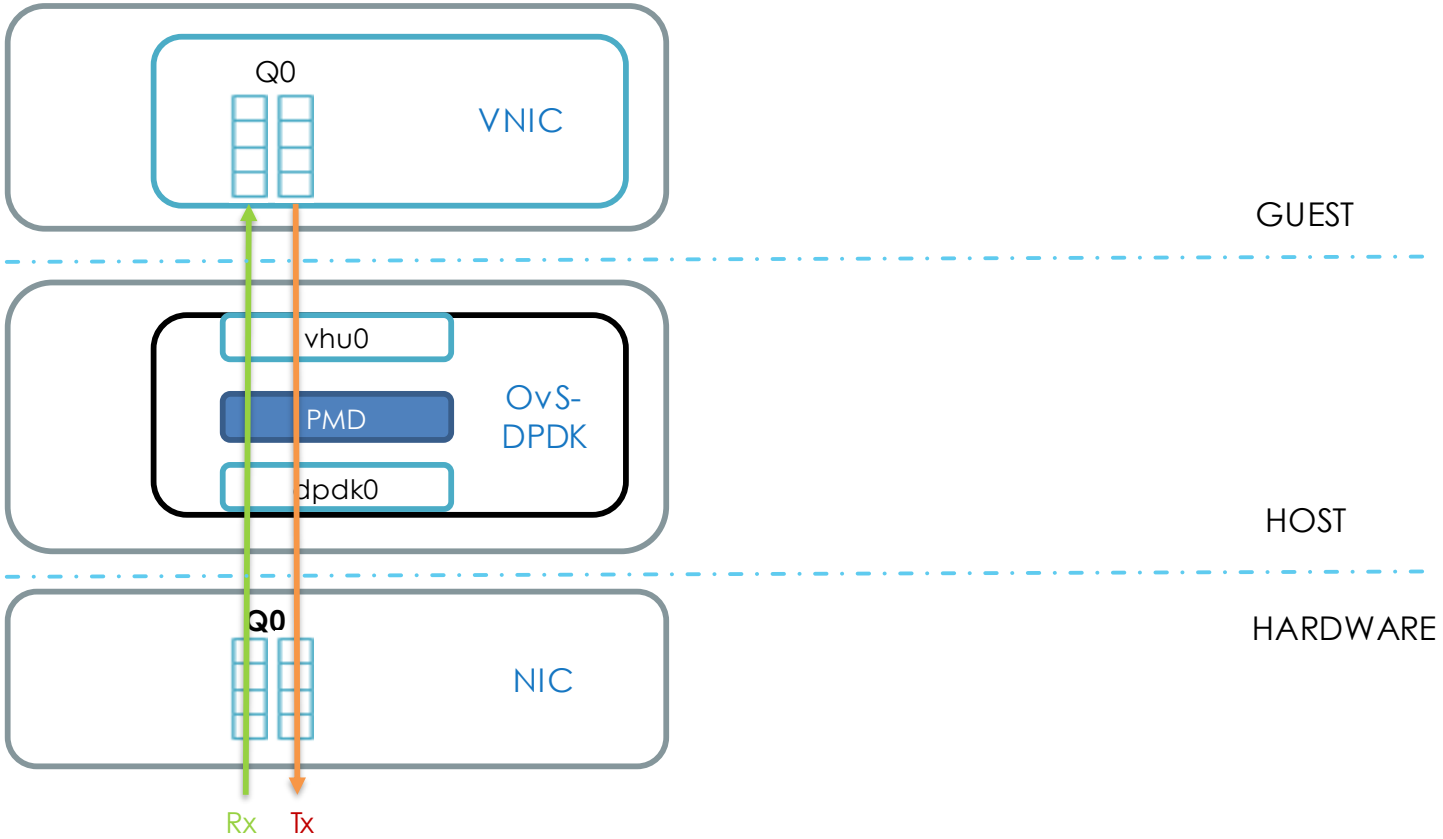


Enable TSO in the guest to reduce vCPU load & boost throughput for OvS-DPDK

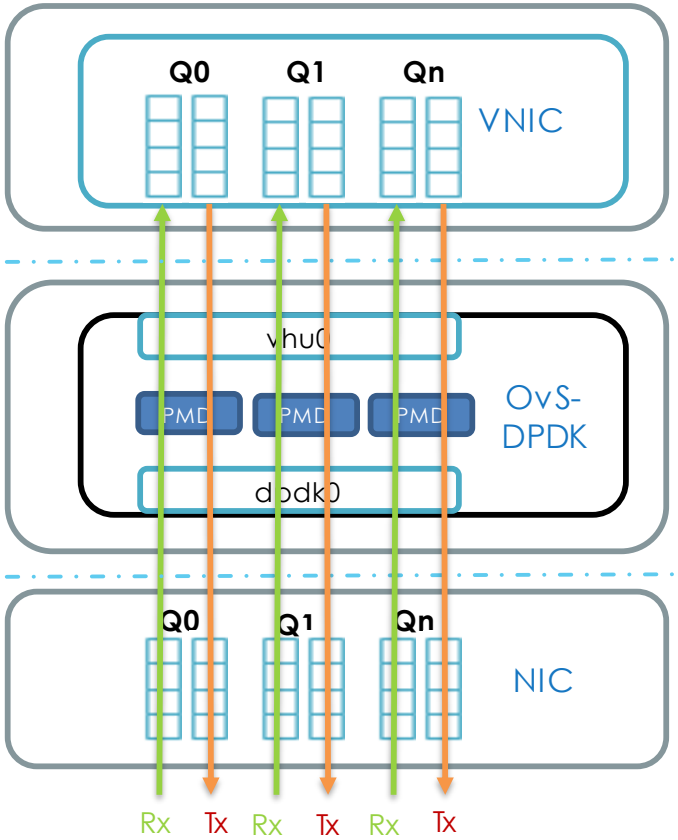
Optimizations: TCP Segmentation Offload



TCP Optimizations: Multi Q (Overview)



TCP Optimizations: Multi Q (Overview)



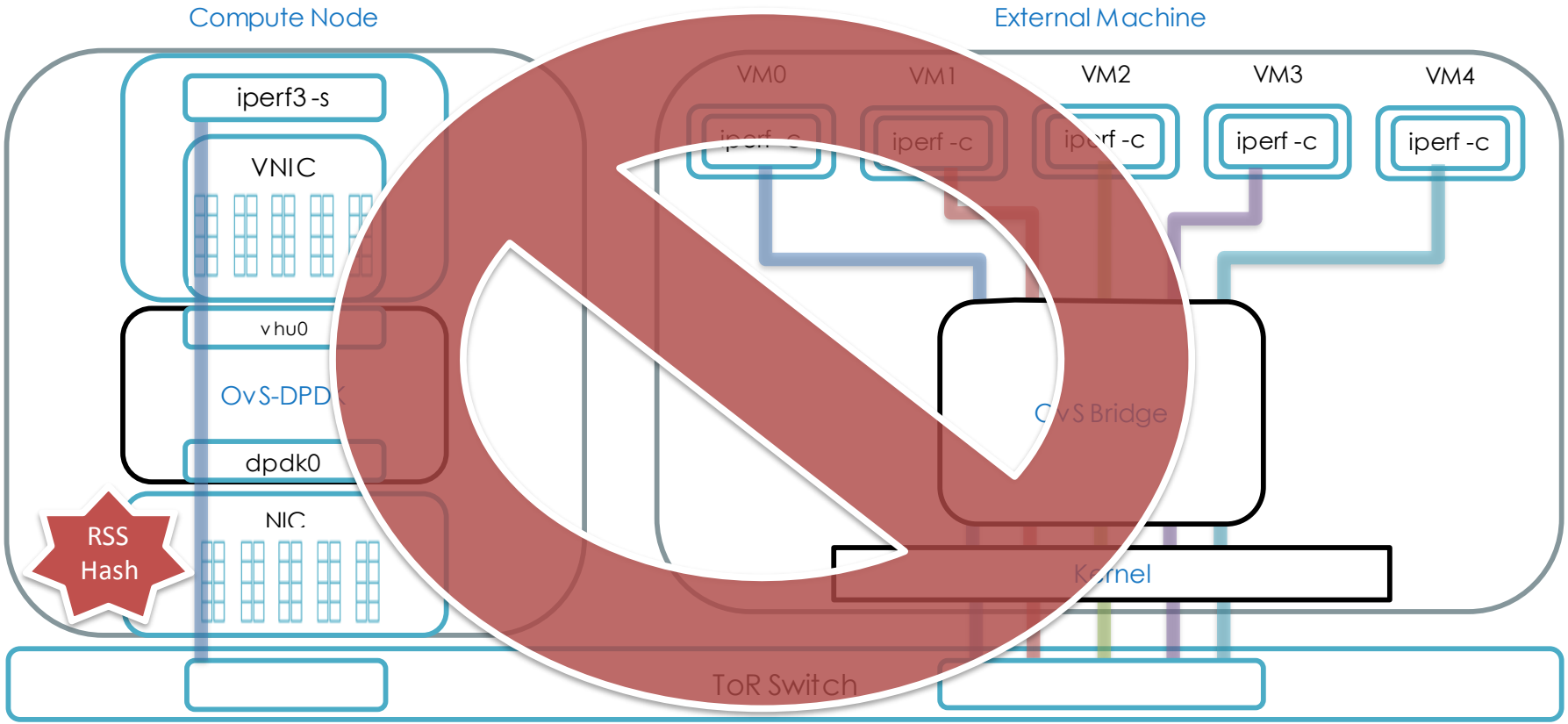
GUEST

HOST

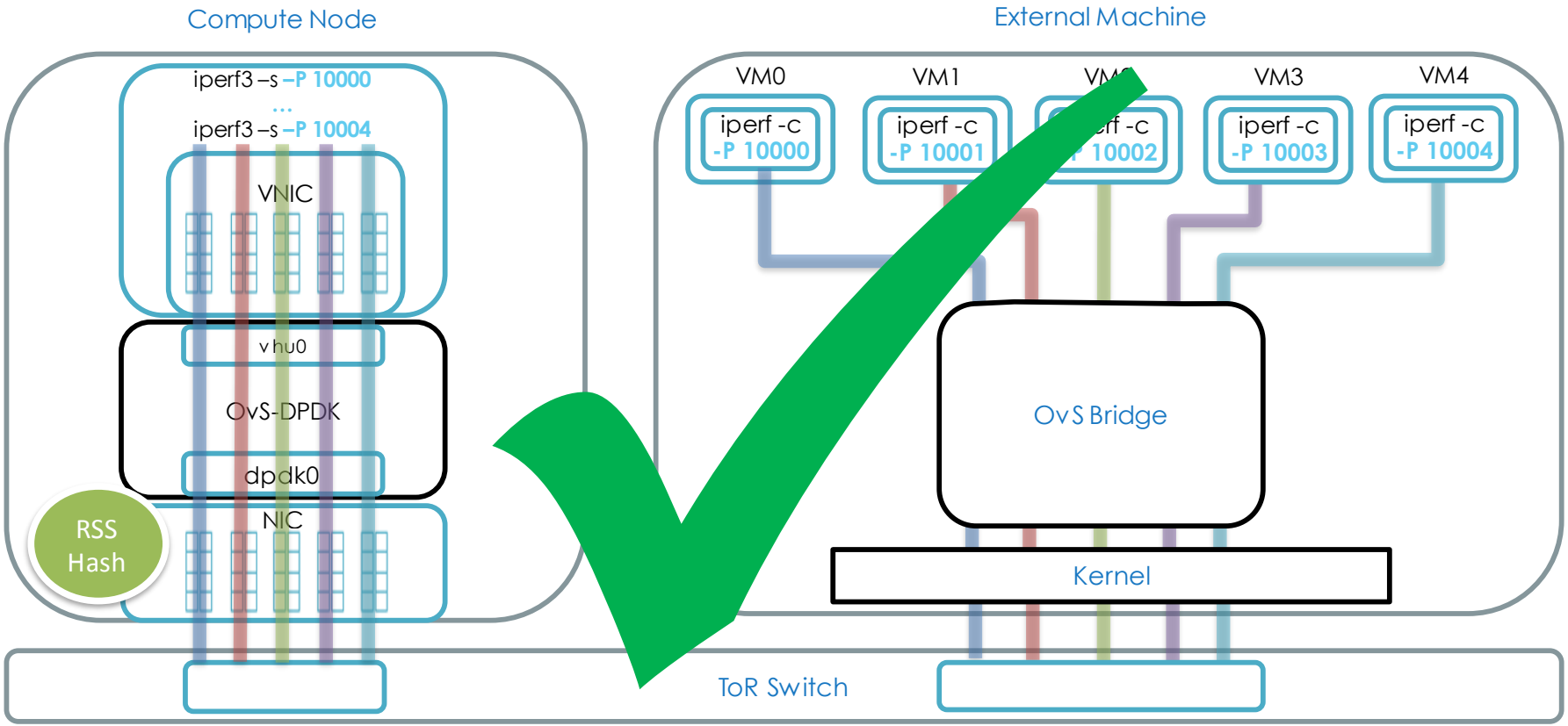
HARDWARE

https://software.intel.com/en-us/articles/configure-vhost-user-multiqueue-for-ovs-with-dpdk

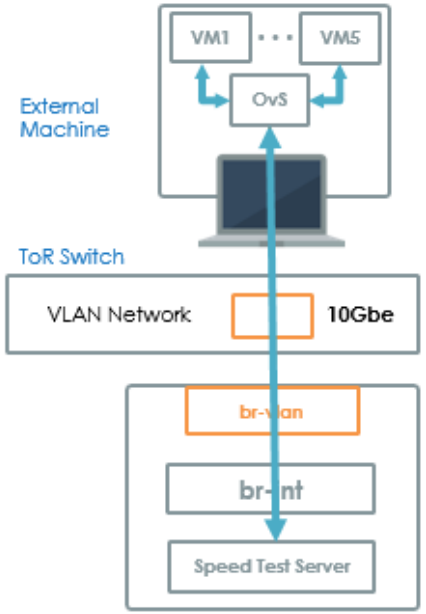
TCP Optimizations: Multi Q (Problem)



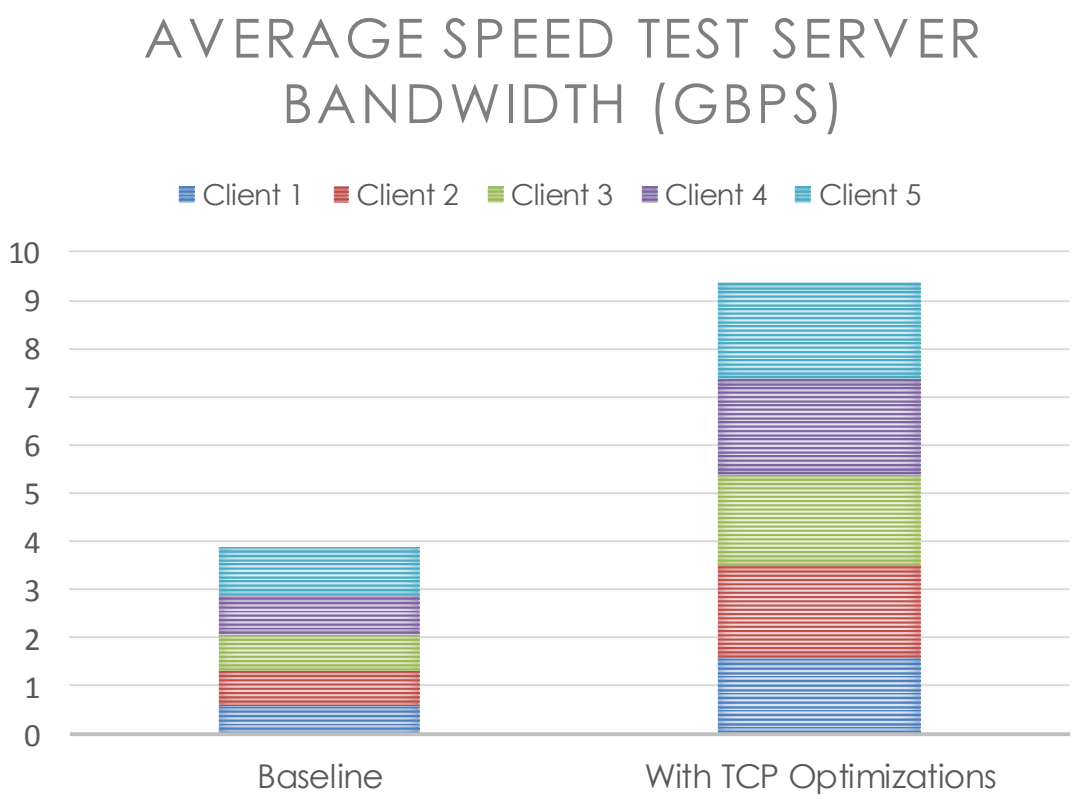
TCP Optimizations: Multi Q (Solution)



Performance Results – Test Case #1

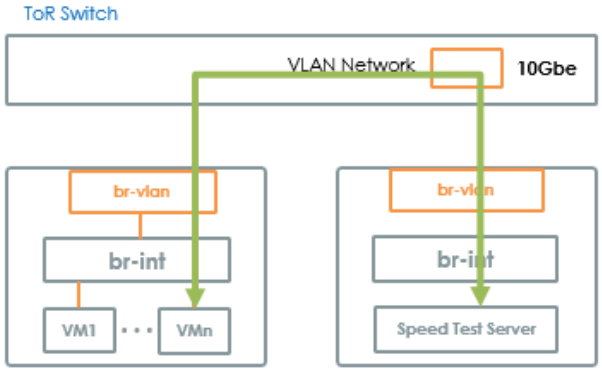


5 x EXTERNAL VM -> SINGLE SPEED TEST SERVER

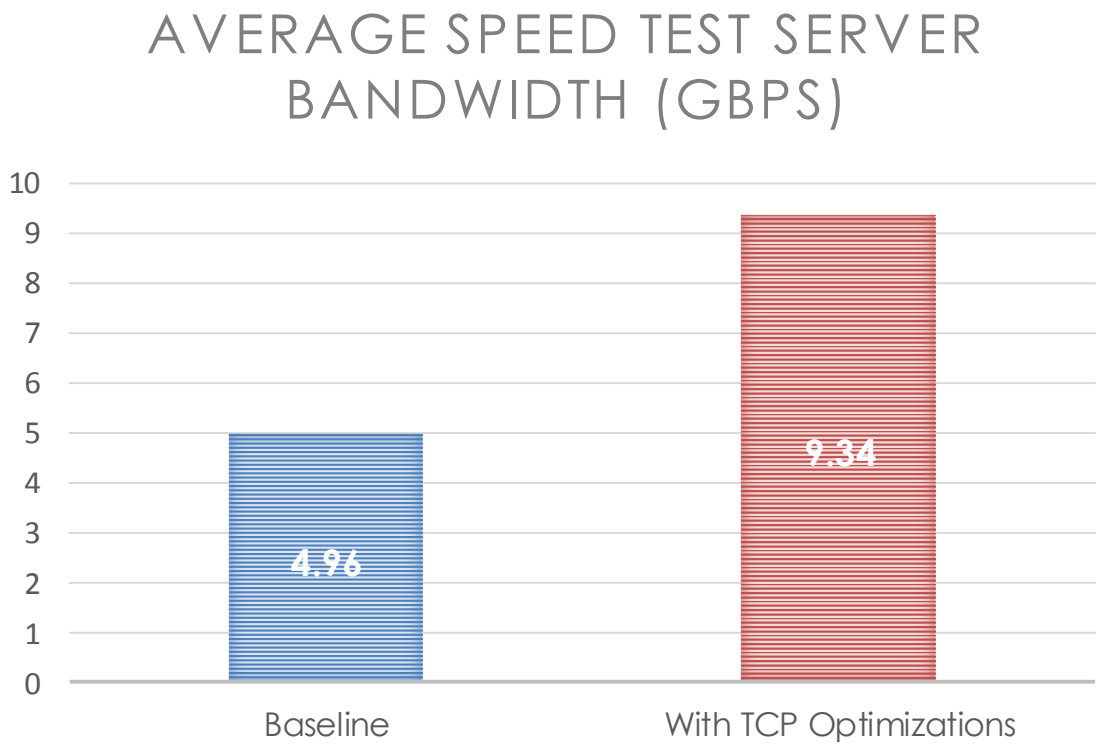


*System configuration detailed in backup

Performance Results – Test Case #2

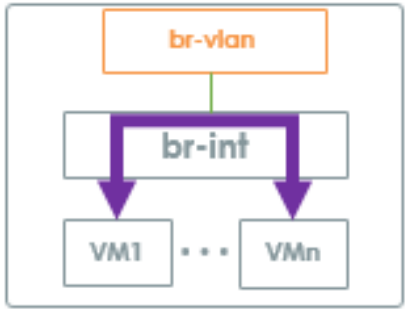


SPEED TEST SERVER -> VM
- SEPARATE COMPUTE NODES -

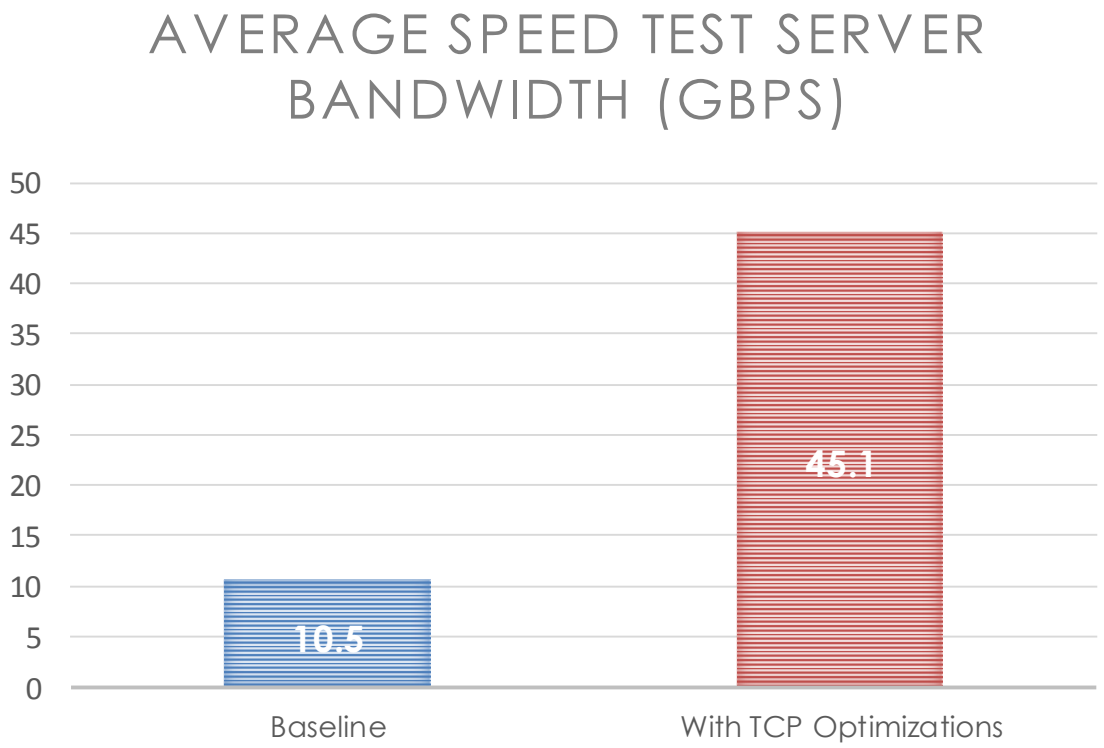


*System configuration detailed in backup

Performance Results – Test Case #3



VM -> VM
SAME COMPUTE NODE



*System configuration detailed in backup

Optimization Summary

Baseline Optimizations

- Enable hugepages
- Per-port/RxQ PMD
- Affinitize workloads
- Incorporate NUMA support

Avail of Offloads

- TSO = reduced vCPU load
- TSO = efficient PCI bandwidth consumption

Utilize Multi Q for Guests

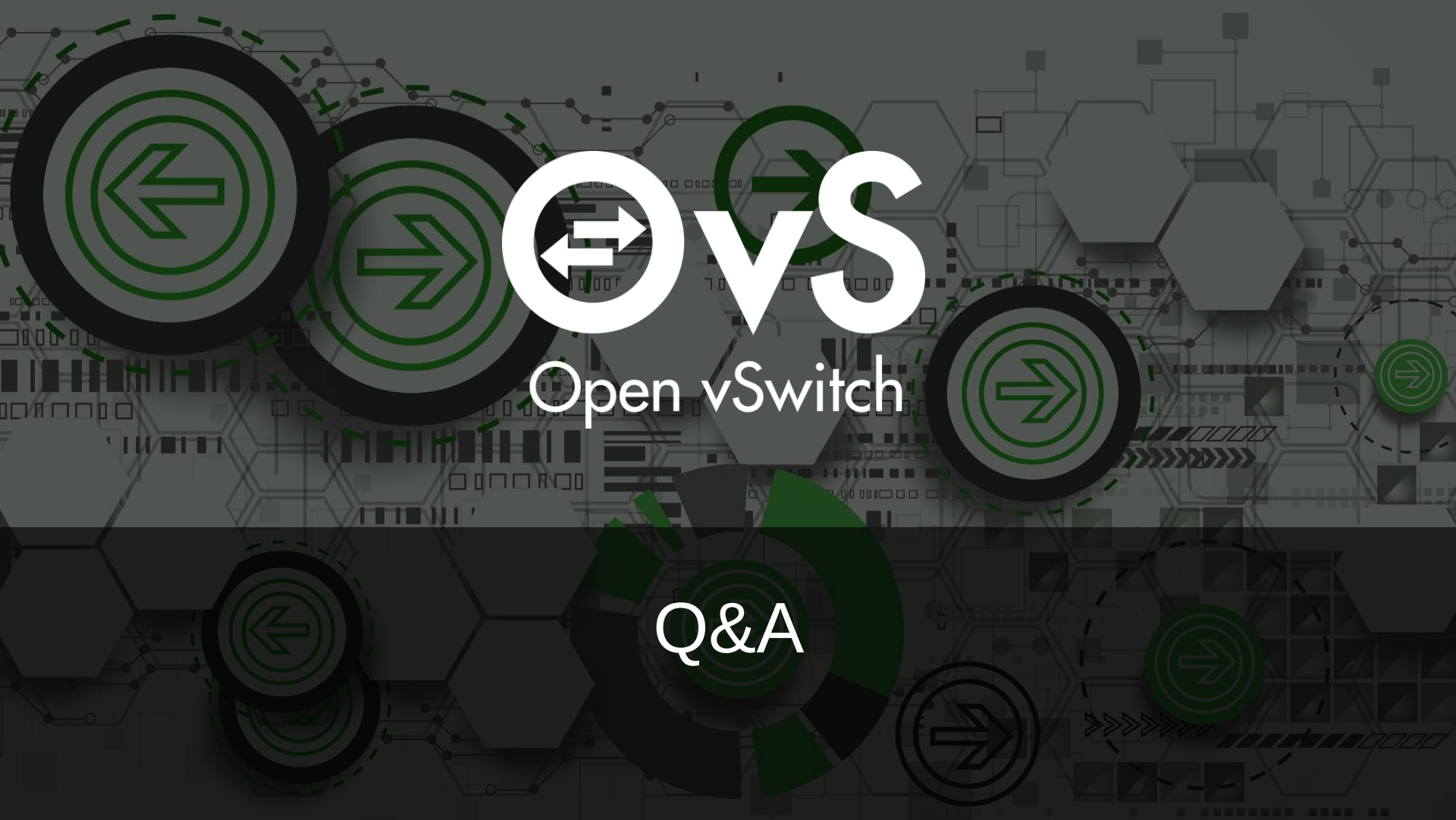
- Saturate line
- Push bottleneck back to the network

Next Steps

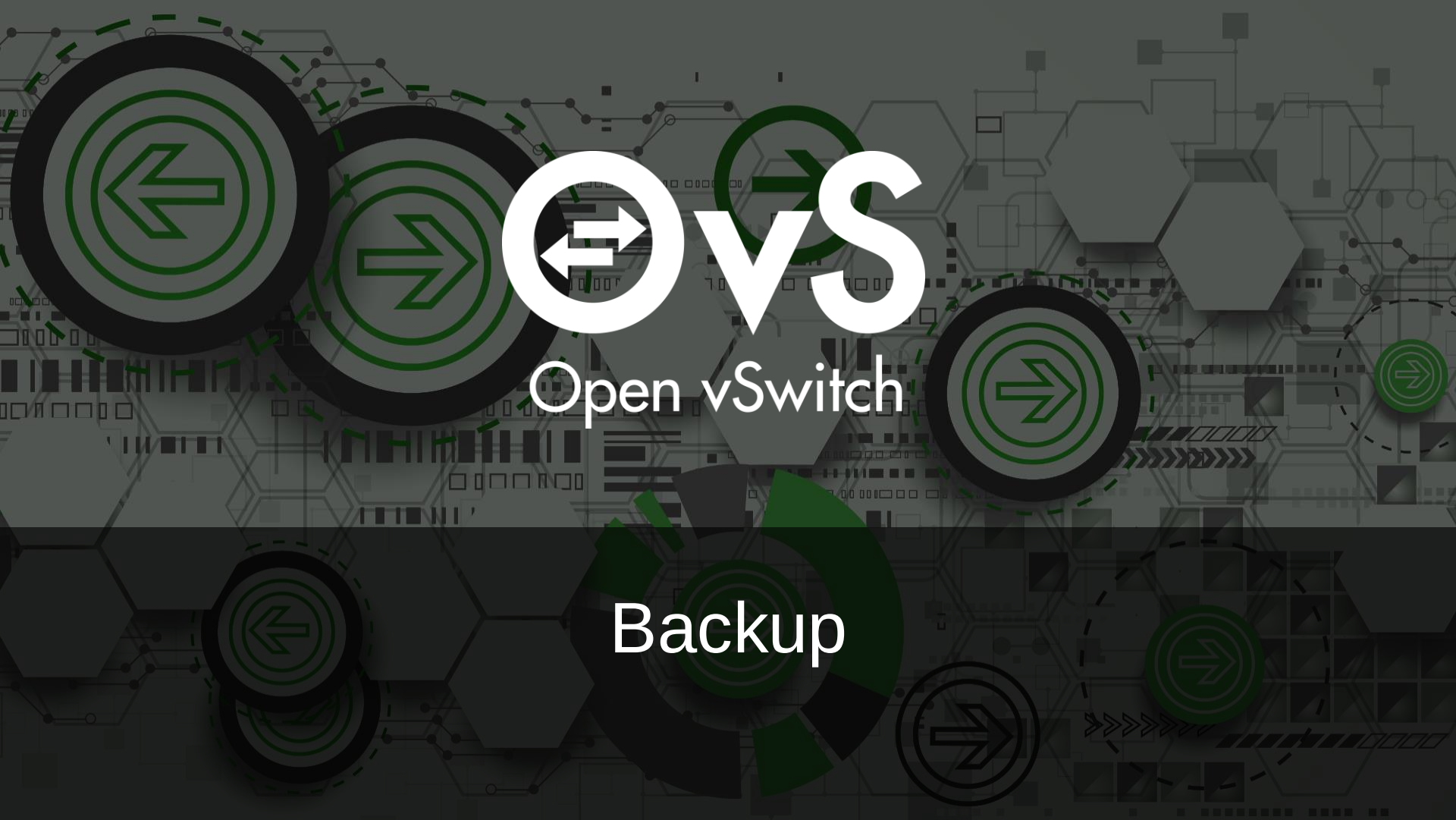
- Release non-RFC TSO Support Patch
- Add support for TSO + Tunnels

References

- https://www.measurementlab.net/publications/understanding-broadband-speed-measurements.pdf
- https://software.intel.com/en-us/articles/configure-vhost-user-multiqueue-for-ovs-with-dpdk
- http://openvswitch.org/pipermail/dev/2016-June/072871.html



Q&A



Backup

System Configuration: Hardware

Hardware Platform Specification

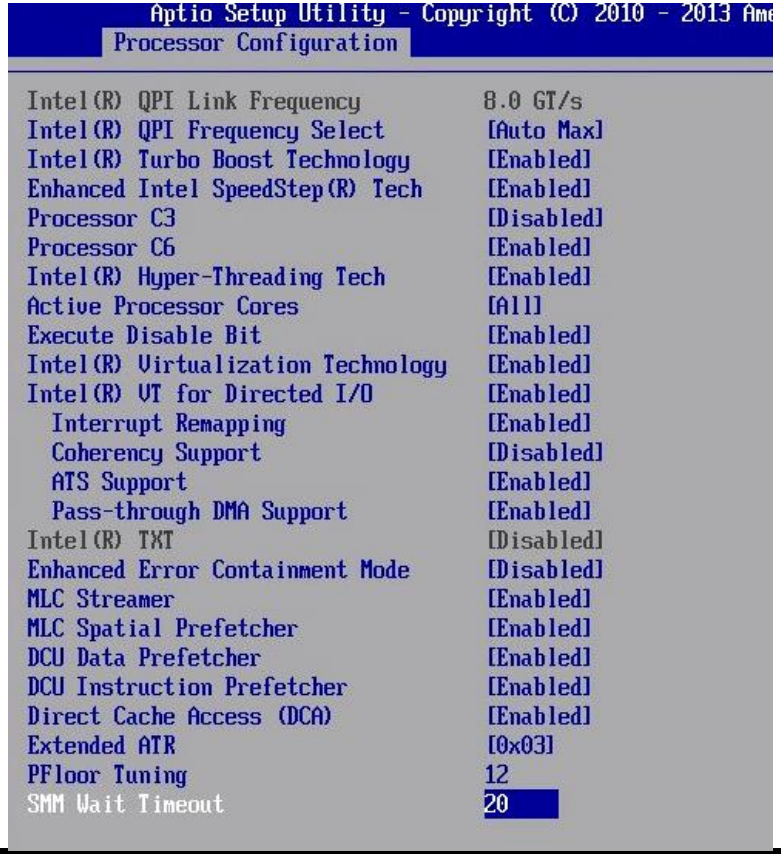
Server	Processor	Hard Drive	Memory	NIC
Compute1	Intel® Xeon® E5-2680 v2 at 2.80 GHz, 40 logical cores	1 TB	DDR3 1600 MHz	• Intel Ethernet Controller I350 BT2 (management and public networks) • Intel® 82599 ES-10 Gigabit Ethernet Controller (VxLAN and VLAN networks
Compute2	Intel® Xeon® E5-2680 v2 at 2.80 GHz, 40 logical cores	1 TB	DDR3 1600 MHz	• Intel Ethernet Controller I350 BT2 (management and public networks) • Intel® 82599 ES-10 Gigabit Ethernet Controller (VxLAN and VLAN networks

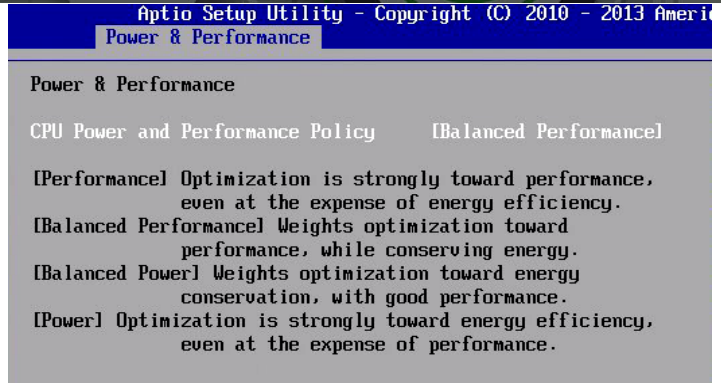
System Configuration: Software

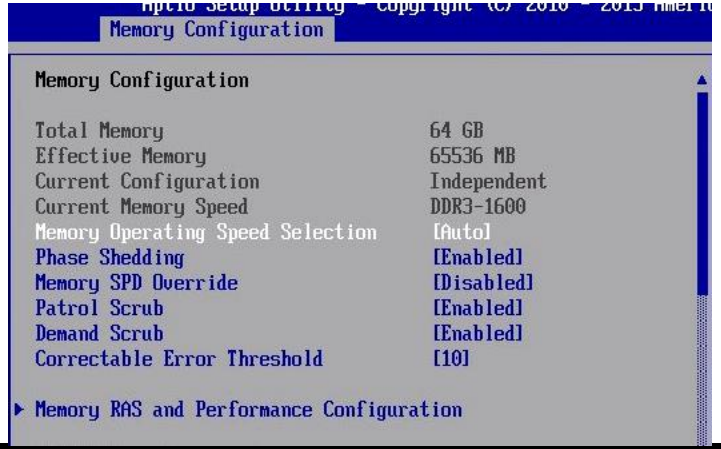
Software Ingredients

#	Software BOM Item	Component
1	Operating System	Fedora* 21, Kernel 4.1.13-100.fc21.x86_64
2	Hypervisor	Compute nodes: QEMU-KVM, QEMU 2.5.0
3	Virtual Switch	Compute nodes: Open vSwitch 2.5.9 + TSO RFC patch
4	Packet Processing Acceleration	DPDK v16.04
5	Virtualized Infrastructure Manager	OpenStack* Kilo 2015.1.0

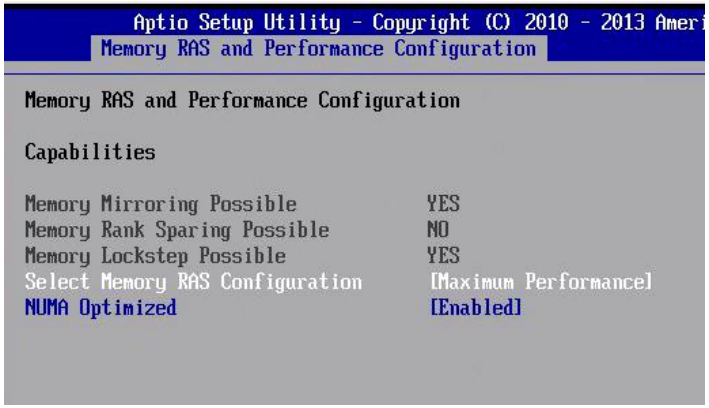
System Configuration: BIOS Settings 1/2

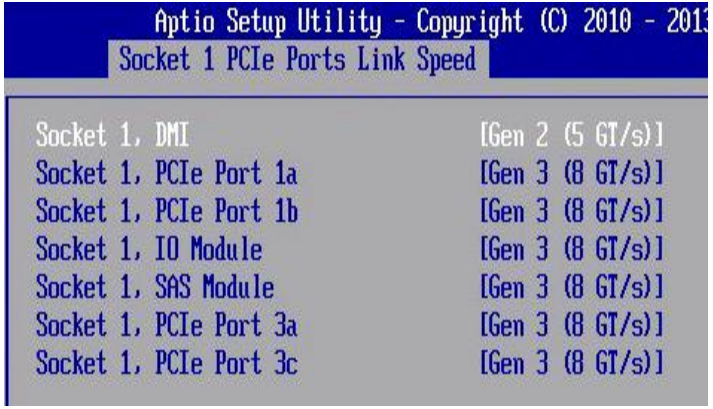


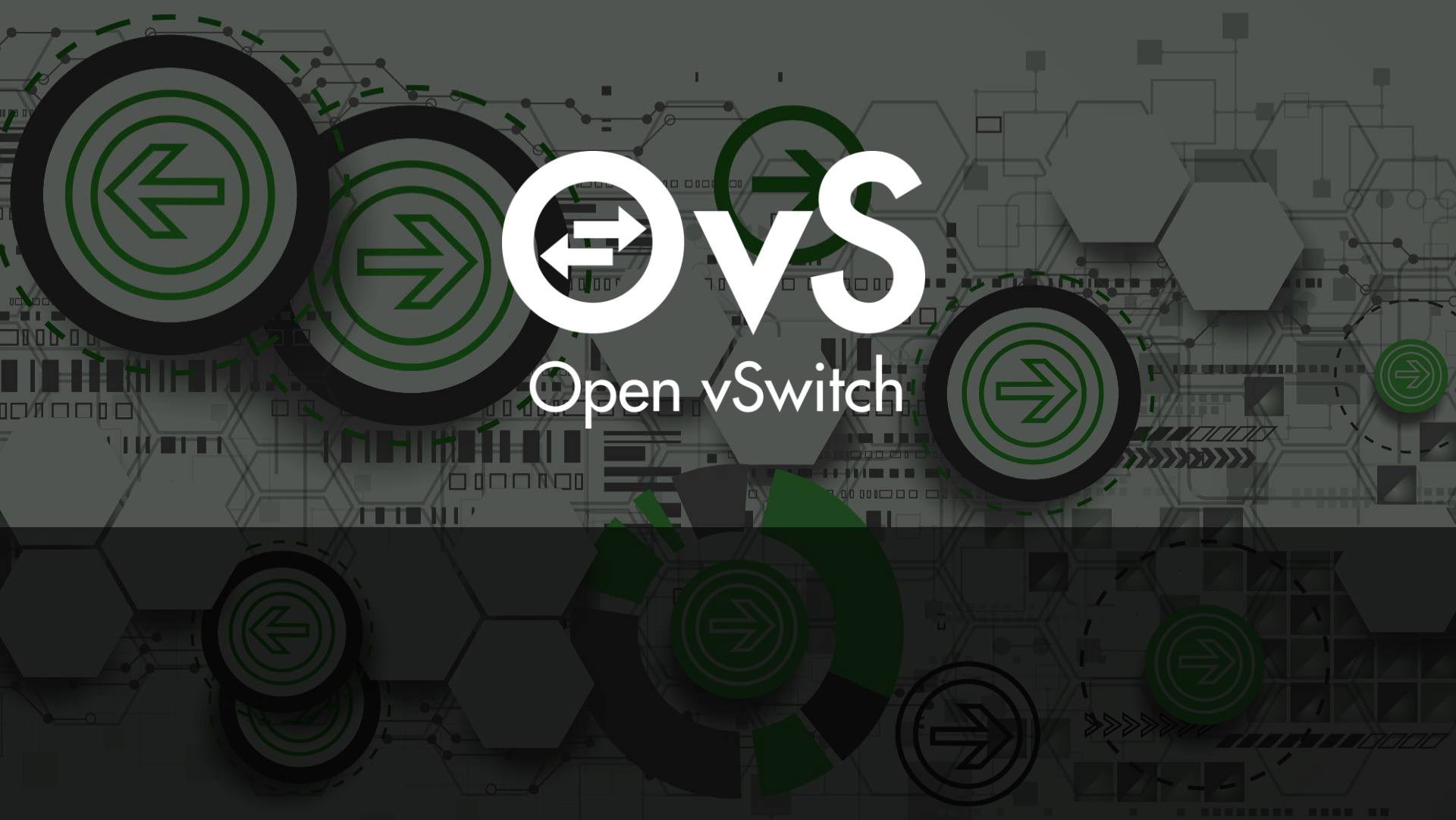




System Configuration: BIOS Settings 2/2






OvS
Open vSwitch