Evaluation and Characterization of NFV Infrastructure Solutions on HP Server Platforms

DPDK Summit

Al Sanders

August 17, 2015

Agenda

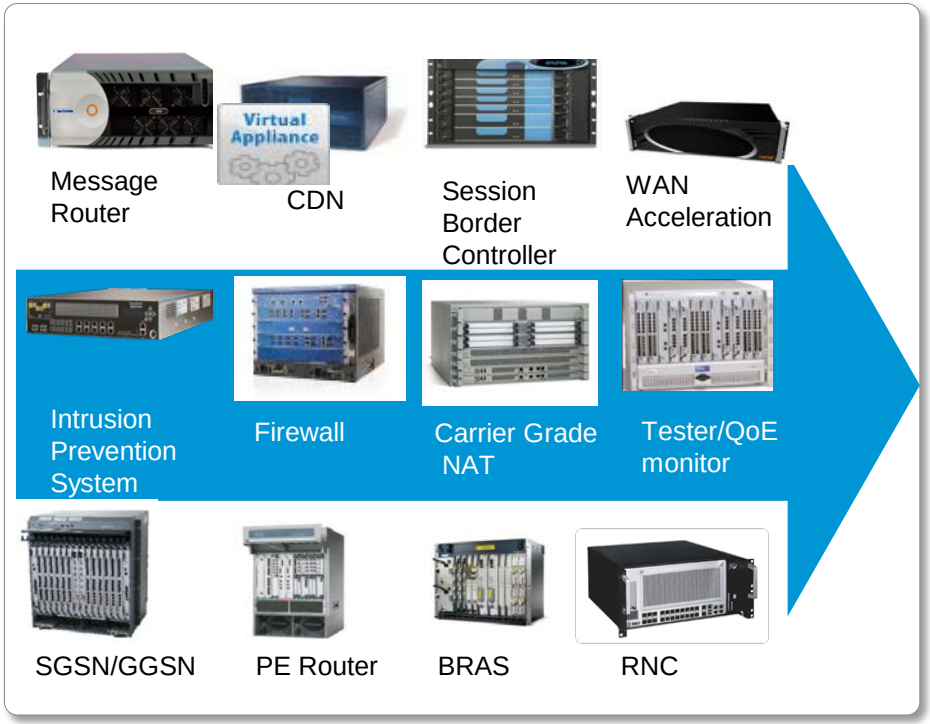
- HP’s Packet Processing Initiative
- DPDK Test Plans and Configurations
- DPDK Test results with Intel ONP
- 6WIND DPDK-based application benchmarks
- Areas for Future Study
- Conclusion

HP’s Packet Processing Initiative

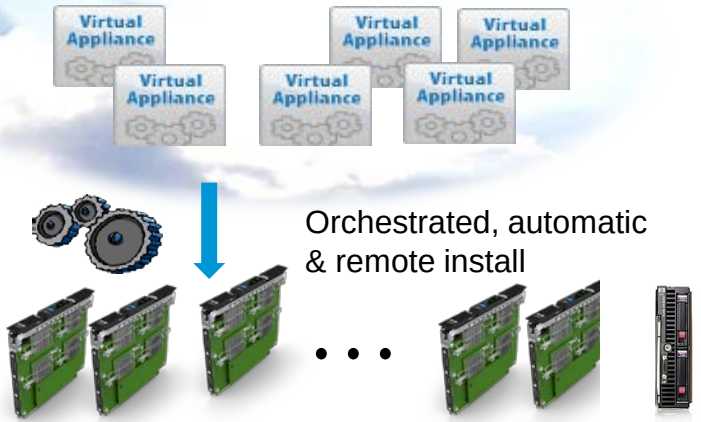
Telecom Operators are embracing the IT-Cost Model

lower costs, reduced time-to-market, and increased business agility

Proprietary Network Appliances



Software Appliances



HP's High Volume IT Servers for Telecom

The Leading IT Servers for Telecom Operators

Ideal systems for IT, Network Data Center and C.O. Deployments

- -48v, NEBS and ETSI compliant configurations at commercial prices
- Full range of OS's & Hypervisors on carrier-grade systems
- Extended product lifecycle, and lifecycle management for telecom

Ideal systems for NFV Infrastructure

- HP's OpenNFV Reference Architecture platform
- Industry's leading NFV, SDN and OpenStack Platform

Optimized and Characterized for Telecom Data Plane Applications

- Industry leading NICs with SR-IOV
- Data Plane Development Kit (DPDK) Support and Optimization
- Intel Open Network Platform Components

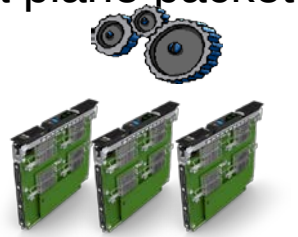
Industry leading Systems Management

- HP OneView, integrated with HP Helion/OpenStack & HP's OpenNFV Orchestrator
- OpenHPI with open source support for Telecom OSS systems

HP Server's Packet Processing Optimization Initiative

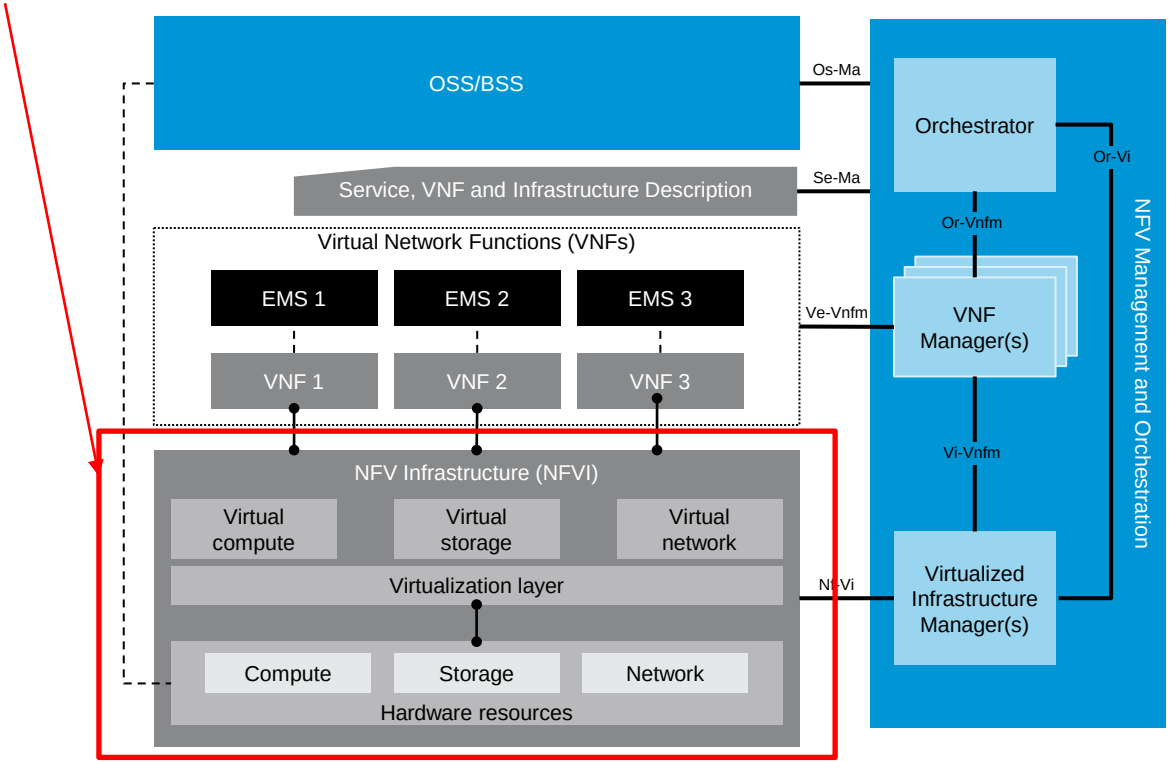
ProLiant the Leading IT Industry Servers for Packet Processing

- HP works very closely with leading enabling technology partners to ensure that HP servers are optimized for the most demanding data plane packet processing applications
- **HP Server's Telecom NFV Infrastructure Lab**
  - Optimize & benchmark packet processing HW & SW
  - Specifically ensure base hardware/software configs enable maximum possible packet processing throughput across the following scope: OS, NICs, Drivers, DPDK, SR-IOV, Hypervisor, vSwitch
  - Interact with OEM partners and infrastructure vendors to resolve functional and performance issues
  - Provide support for selected customer PoCs and demos



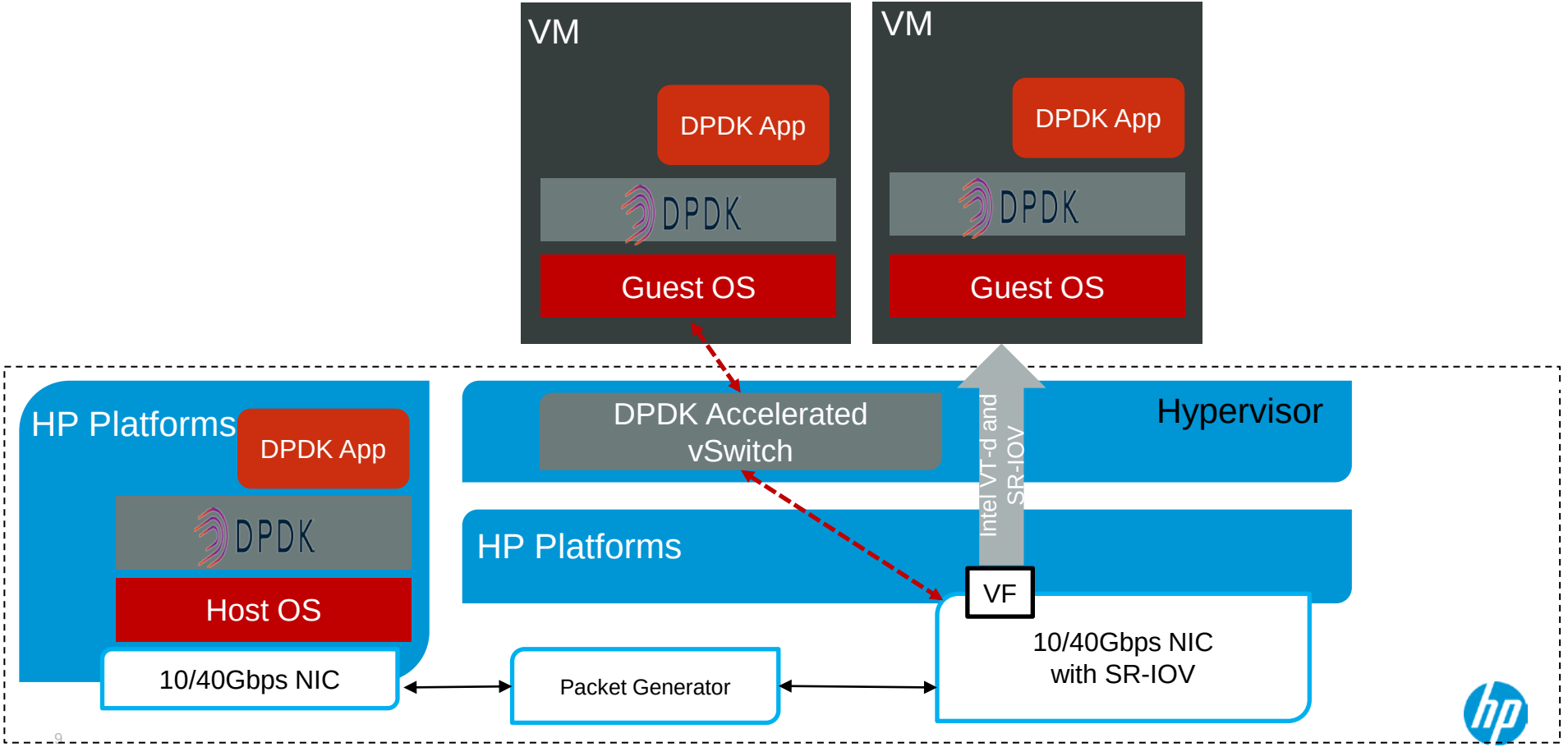
ETSI NFV Architecture Diagram

Packet processing performance is key to NFV Infrastructure effectiveness



DPDK Test Plans and Configurations

DPDK Test Configurations



DPDK Test Cases

Following Intel ONP (Open Network Platform) Server 1.3.1 Benchmark Test Report

1.1	Host L2 Fwd:	DPDK-based port forwarding sample application.
1.2	Host L3 Fwd:	DPDK-based port forwarding sample application.
2.1	Host OVS L2 Fwd :	OvSwitch + DPDK-netdev
2.2	Host OVS L3 Fwd :	OvSwitch + DPDK-netdev
3.1	VM L2 Fwd + SRIOV:	DPDK-based port forwarding with passthrough
3.2	VM L3 Fwd + SRIOV:	DPDK-based port forwarding with passthrough
4.1	VM OVS L3 Fwd :	vSwitch on Host, single VM
4.2	VM OVS L3 Fwd :	vSwitch on Host, Dual VM, SERIES Traffic
4.3	VM OVS L3 Fwd:	vSwitch on Host, Dual VM, PARALLEL Traffic

Current Testing Configuration

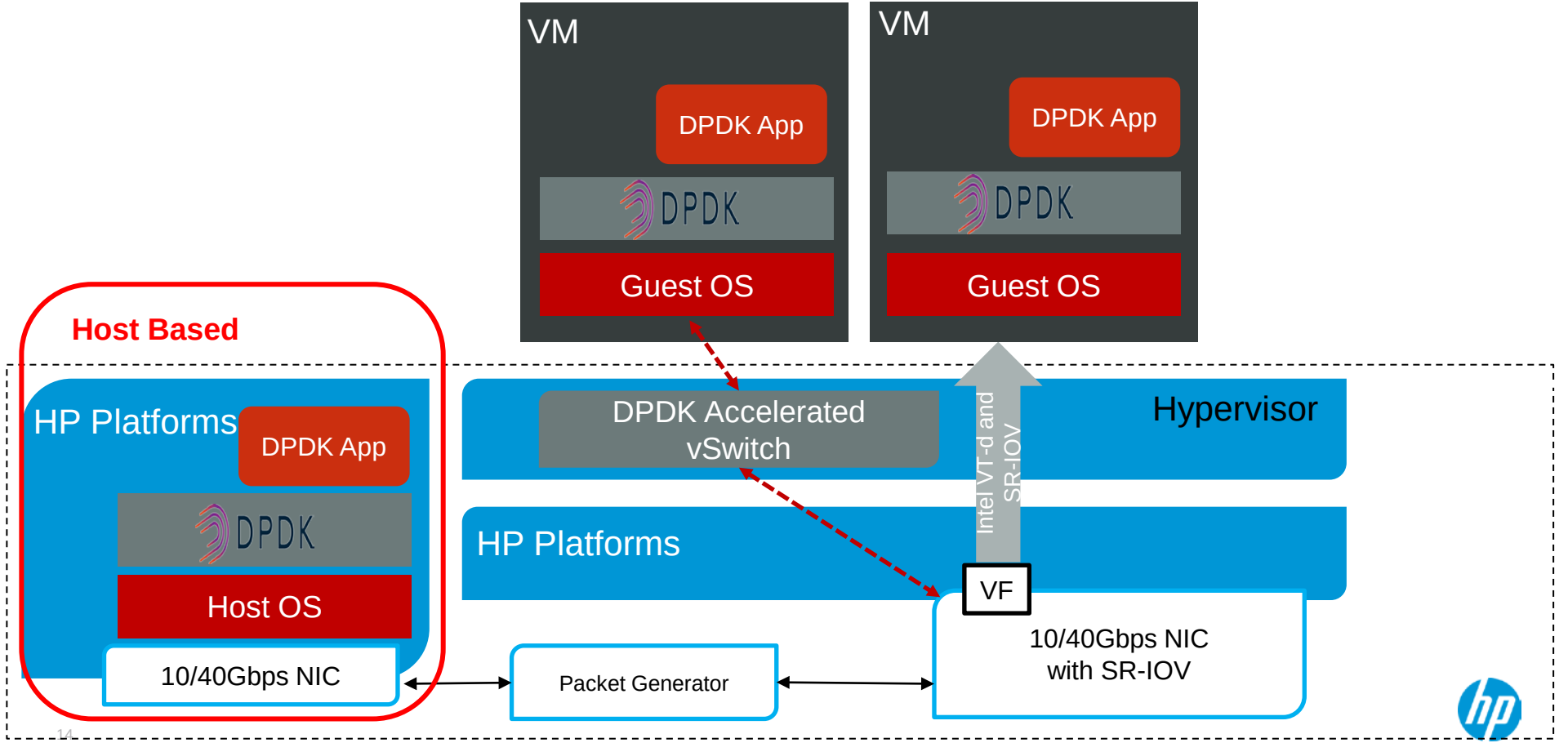
- Location
  - Ft. Collins, CO data center (6 racks)
  - Palo Alto, CA data center (2 racks)
- Server HW
  - DL380 Gen9's rack-mount
  - C7000 with BL460 Gen9 blades
  - Moonshot
- NICs
  - Intel (Niantic, Fortville)
  - QLogic
  - Mellanox
  - Avago/Emulex
- OS/SW
  - RedHat, Wind River, Fedora, Debian, hLinux, 6WIND
- Spirent N11U Network Test Generator
  - 40/10GbE QSFP+ ports

Hardware Network Test Generator essential for DPDK testing

- Dpdk-pktgen used for initial testing
- Formal testing uses Spirent
  - Reliable line rate 1/10/40GbE
  - HW measured packet return timing
  - Capacity up to 5x40GbE
  - Automated Industry Standard tests
    - RFC2544 tests built in.
    - Max zero pkt loss throughput @ each pkt size automatic discovery
    - Accurate Latency tests based on measured max safe throughputs for each packet size
    - Built in data report generation

DPDK Test Results with Intel ONP

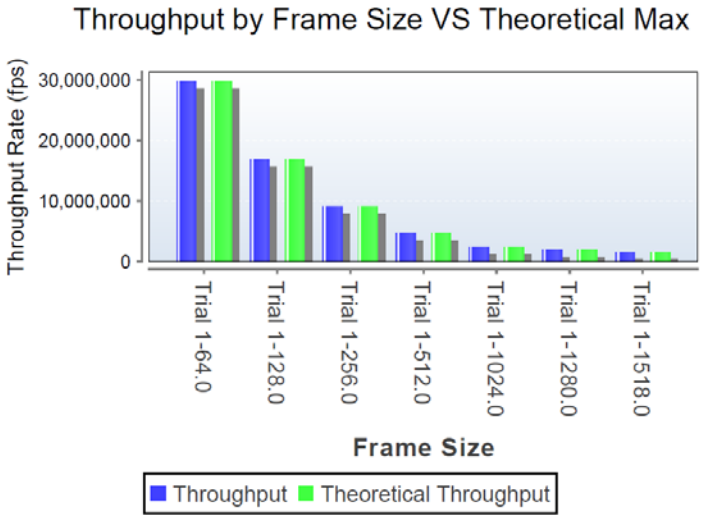
DPDK Test Configurations



Host l2fwd Throughput Example

Intel Fortville X710-DA4 NIC, with i40e PMD

- l2fwd is distributed with the DPDK release
- Bi-directional packet flows through two of the ports
- Processor receives packet on one NIC port, then immediately sends it out the other NIC port.
- No processor time is spent analyzing the packet.
- RFC 2544 measures maximum lossless packet rate
- Intel® Fortville achieves 10 GbE line rate for all packet sizes
    - Had to experiment with initial load number for RFC 2544 test run to achieve 100% load at 64 bytes
- Following tuning guidelines for l2fwd in Appendix E of ONP 1.3.1 Benchmark Test Report resulted in improved results
    - Increase NIC I/O queue sizes

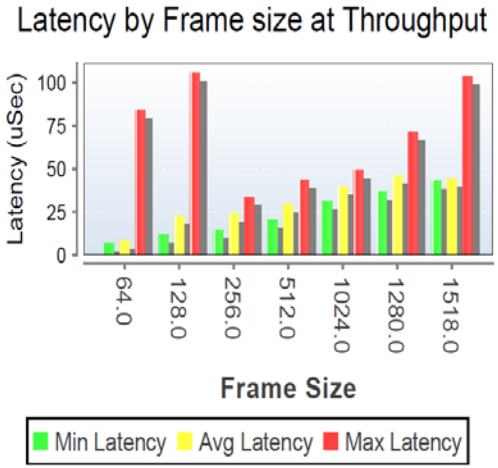


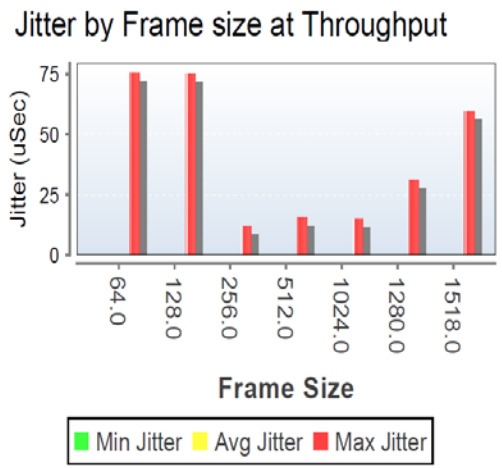
Total Trials	Number of Passed Trials	Frame Size (bytes)	Intended Load (%)	Offered Load (%)	Throughput (%)	Aggregated Throughput (fps)	Aggregated Theoretical Max (fps)	Aggregated Throughput (Mbps)
1	1	64	100	100	100	29761904.767	29761904.762	20000
1	1	128	100	100	100	16891891.9	16891891.892	20000
1	1	256	100	100	100	9057971.017	9057971.014	20000
1	1	512	100	100	100	4699248.133	4699248.12	20000
1	1	1024	100	100	100	2394636.017	2394636.015	20000
1	1	1280	100	100	100	1923076.933	1923076.923	20000
1	1	1518	100	99.87	99.87	1623376.633	1625487.646	19974.026

Host l2fwd Latency Example

Intel Fortville X710-DA4 NIC, with i40e PMD

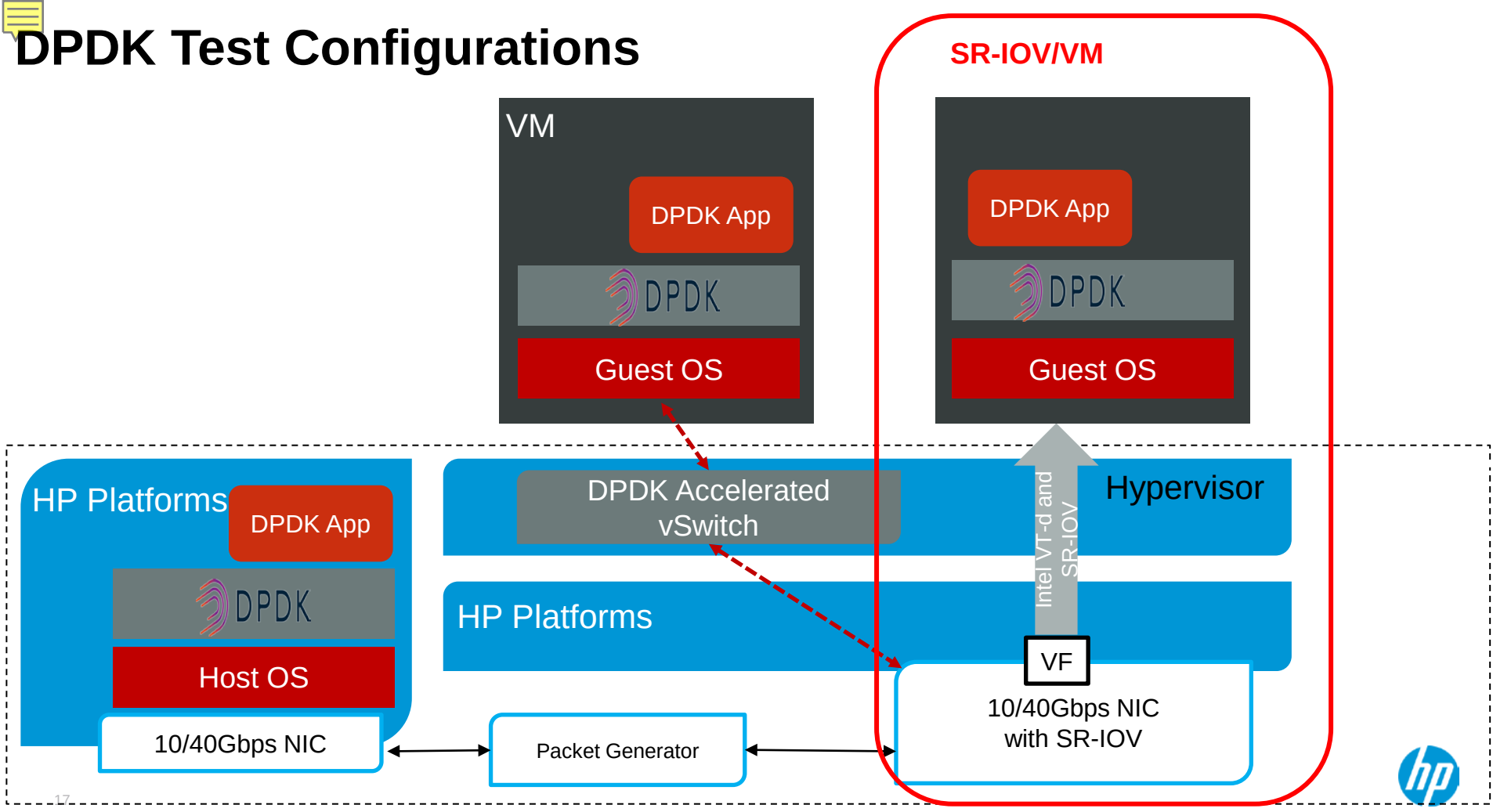
- Latency measurement shows time delta from last bit of transmitted packet to last bit of received packet
- Average latency increases linearly with packet size
- Latency numbers show 10-20% improvement over Niantic
- Table shows binary search algorithm to obtain highest load with zero packet loss
- Variability in maximum latency numbers remains an area of investigation
- NW latency vs. Interrupt latency
  - Primarily a function of HW
- Jitter measures latency variance





Frame Size (bytes)	Intended Load (%)	Offered Load (%)	Min Frame Loss (%)	Min Latency (uSec)	Avg Latency (uSec)	Max Latency (uSec)	Min Jitter (uSec)	Avg Jitter (uSec)	Max Jitter (uSec)
64	98	95.455	0.000472325332930282	6.88	8.329	106.405	0	0.013	98.3
64	49.5	48.837	0	5.93	8.821	49.035	0	0.264	40.91
64	73.75	72.414	0	6.26	8.455	94.295	0	0.096	86.46
64	85.875	84	0	6.453	7.707	94.45	0	0.048	87.38
64	91.938	91.304	0	6.883	8.25	16.763	0	0.024	38.36

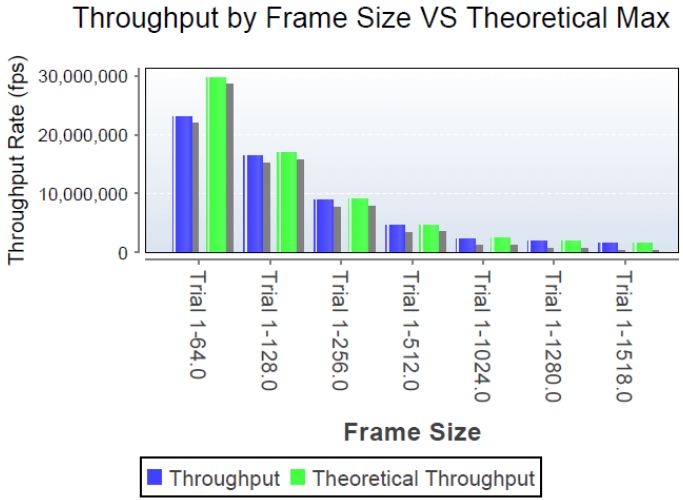
DPDK Test Configurations



VM L2 Fwd + SRIOV Throughput Example

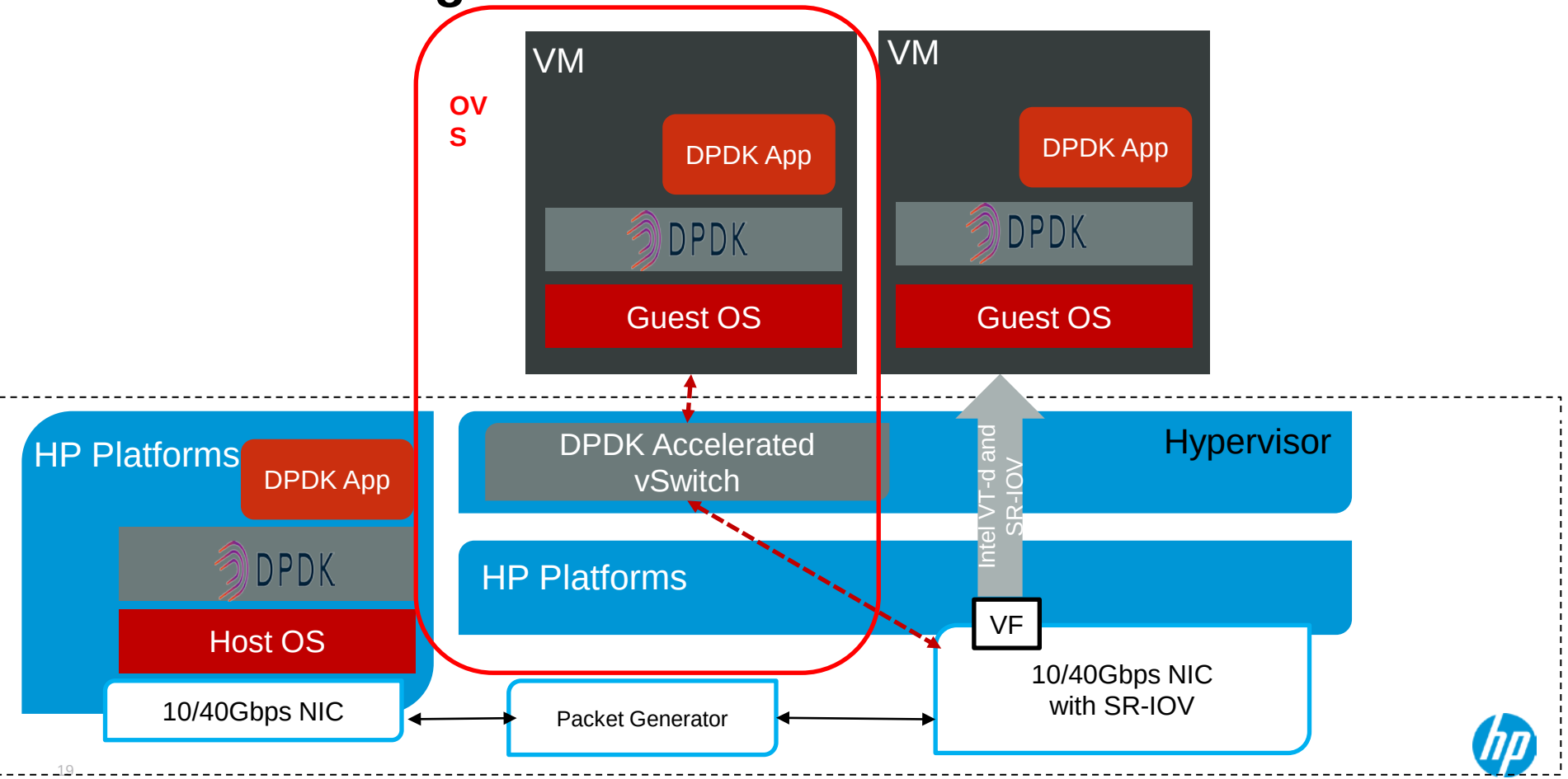
HP 560SFP+ 10GbE 2-port adapter (Niantic)

- SR-IOV with PCI passthrough of virtual functions.
- Performance for SR-IOV in VM nearly identical to bare-metal performance in host OS.
- Near line rate performance for packet sizes >64 bytes
- Need to set hugepages to 1Gb to reduce packet loss for small packets
  - RHEL 7 implements 1GB hugepages for VM



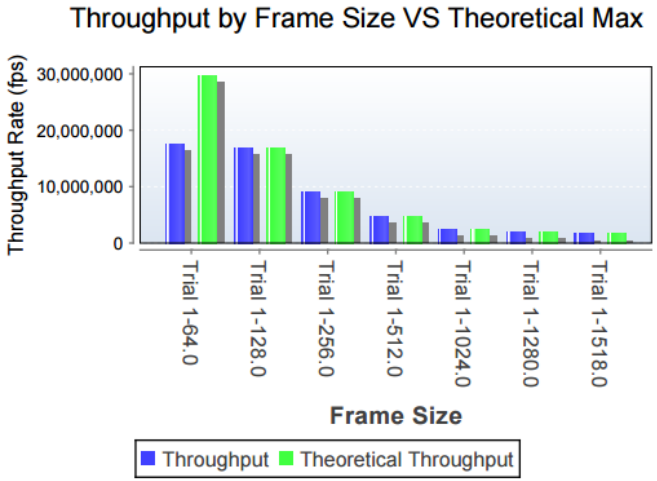
Total Trials	Number of Passed Trials	Frame Size (bytes)	Intended Load (%)	Offered Load (%)	Throughput (%)	Aggregated Throughput (fps)	Aggregated Theoretical Max (fps)	Aggregated Throughput (Mbps)
1	1	64	79.688	77.778	77.778	23148148	29761904.762	15555.555
1	1	128	99.219	97.368	97.368	16447368	16891891.892	19473.684
1	1	256	99.219	98.571	98.571	8928570	9057971.014	19714.283
1	1	512	99.219	99.254	99.254	4664178	4699248.12	19850.742
1	1	1024	99.219	99.239	99.239	2376424	2394636.015	19847.893
1	1	1280	99.219	99.085	99.085	1905486	1923076.923	19817.054
1	1	1518	100	99.87	99.87	1623376	1625487.646	19974.018

DPDK Test Configurations



Host OVS results

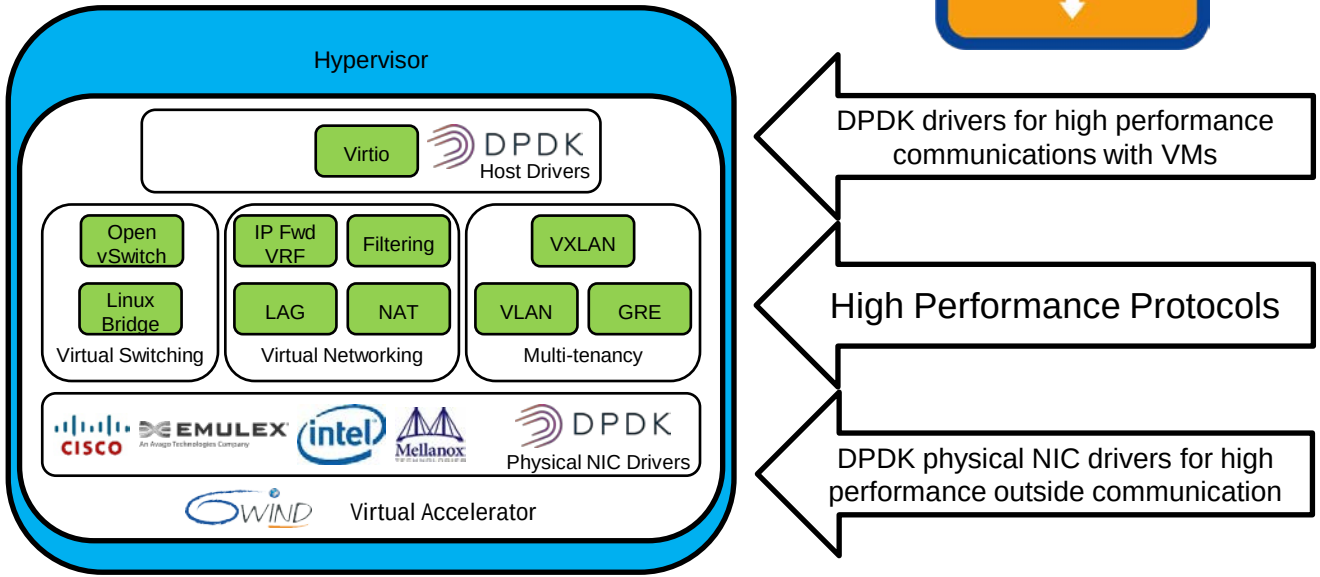
- This test measures raw OVS-dpdk performance
- OVS acts in place of l2fwd application
- Small packet performance is significantly worse (only 58% load achieved at 64 byte packet instead of 78% with Niantic)
- However, other packet sizes achieve near line rate
- This demonstrates that accelerating OVS with DPDK offers significantly improved performance over standard OVS (2Mpps at 64b in one study)



Total Trials	Number of Passed Trials	Frame Size (bytes)	Intended Load (%)	Offered Load (%)	Throughput (%)	Aggregated Throughput (fps)	Aggregated Theoretical Max (fps)	Aggregated Throughput (Mbps)
1	1	64	59.727	58.921	58.921	17536013.067	29761904.762	11784.201
1	1	128	99.922	99.688	99.688	16839228.173	16891891.892	19937.646
1	1	256	99.922	99.688	99.688	9029731.053	9057971.014	19937.646
1	1	512	99.922	99.688	99.688	4684597.313	4699248.12	19937.646
1	1	1024	99.922	99.688	99.688	2387170.28	2394636.015	19937.646
1	1	1280	99.922	99.693	99.693	1917177.92	1923076.923	19938.65
1	1	1518	100	99.87	99.87	1623376.627	1625487.646	19974.026

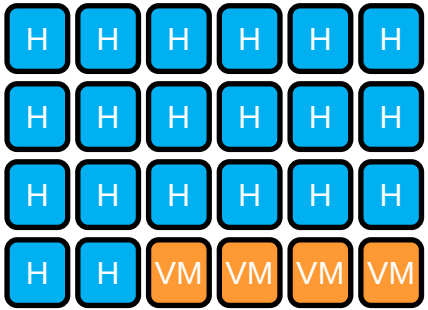
6WIND DPDK-based Application Benchmarks

DPDK Enabled NFV
6WIND Virtual Accelerator

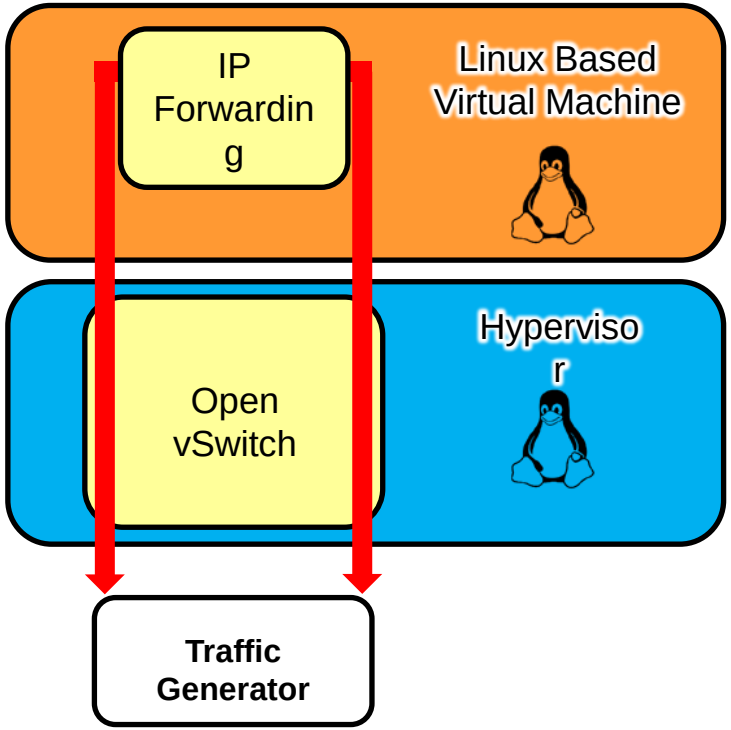


Linux Router + Linux Open vSwitch

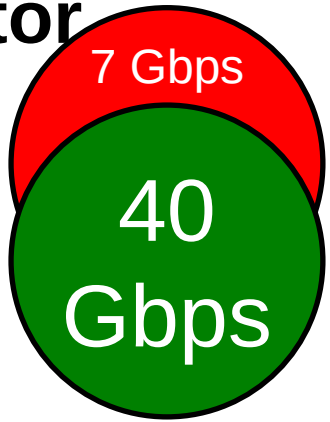
7 Gbps

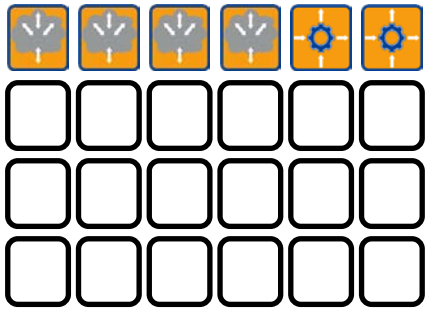


80% of the cores used by hypervisor
➔ VM performance is limited

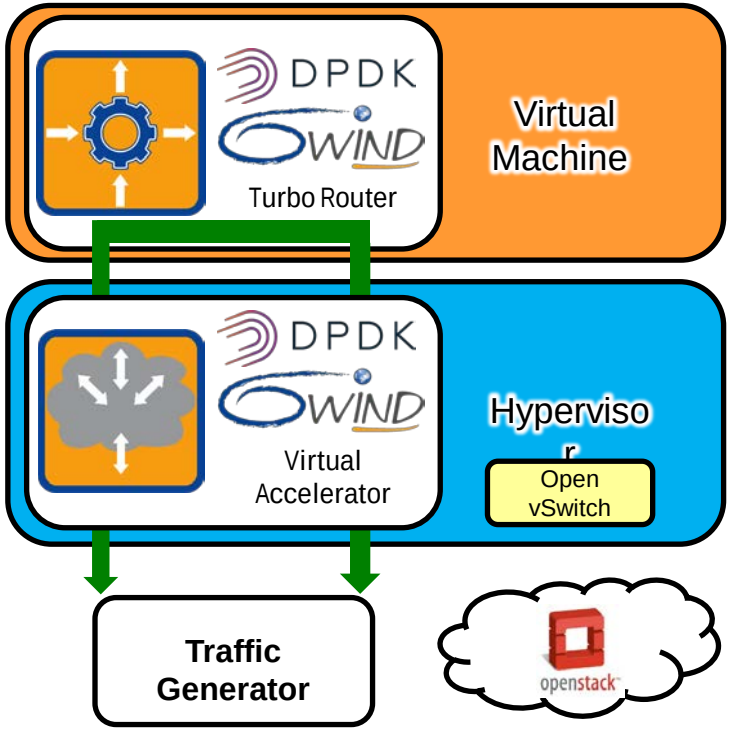


DPDK + 6WIND Turbo Router + Virtual Accelerator





5x more performance; 75% capacity available



Areas for Future Study

Additional OVS/VM testing configs

- Measuring packet performance between multiple virtual machines through the vSwitch.
- This will represent a more realistic NFV application configuration and will indicate how much packet-processing bandwidth can be expected in typical NFV architecture deployments.
- Maintains full VM flexibility where required
  - SR-IOV config restricts migration
- Testing configs
  - Routed through OVS-dpdk to a single VM
  - Routed through OVS-dpdk to dual VMs in series
  - Routed through OVS-dpdk to dual VMs in parallel
- Will use DPDK 2.0 and vhost-user support recently added to OVS netdev-dpdk

Other Areas of Study

- **Expand DPDK test suites beyond l2fwd/l3fwd**
  - Investigate typical VNFs and the load requirements they place on NFVI.
- **Higher speed NICs**
  - 20GbE blade LOMs and Mezz cards
  - 40GbE NICs
  - 25/50/100 GbE NICs
  - Investigate other bottlenecks, such as PCI bandwidth limitations (ie, 40GbE maxes out PCI Gen3 x8)
- **Investigation of latency results**.
  - Currently, average latency looks reasonable (in the <10 microsecond range), but there is some variability in maximum latency numbers that warrant further study.
- **Different Intel Xeon Processors.**
  - Testing is planned to determine the impact of different Intel Xeon Processor types. Allocation of VMs to separate NUMA zones on the processor should be a promising area of investigation.
- **Support strategies for DPDK versions**
  - Platform vendors will be shipping drivers for a certain version of DPDK. As customer issues are reported, will defect fixes be available for previously released versions? Current top-of-tree support model is not ideal for commercial applications. Need bundled solutions which bring together consistent versions of DPDK, OVS, PMDs, qemu, etc.

Conclusion

- HP's NFV Infrastructure lab focusing on demonstrating viable NFVI performance using DPDK with a broad range of HW/SW configurations on HP Server Platforms
- Results show DPDK technology is enabling packet processing performance to support the most demanding NFV data-plane requirements
- Network vendor support for DPDK is expanding rapidly
  - Encouraging dpdk.org as primary PMD distribution model
- We're committed to working with HW and SW partners in the DPDK/NFV community to promote broad industry acceptance of DPDK