OvS

Open vSwitch

December 10-11, 2019 | Westford, MA

Partial Offload Optimization and Performance on Intel Ethernet 700 Series NICs Using rte_flow

Irene Liew and Chenmin Sun (Intel)

(*) Presentation covered by Mesut Ali Ergin

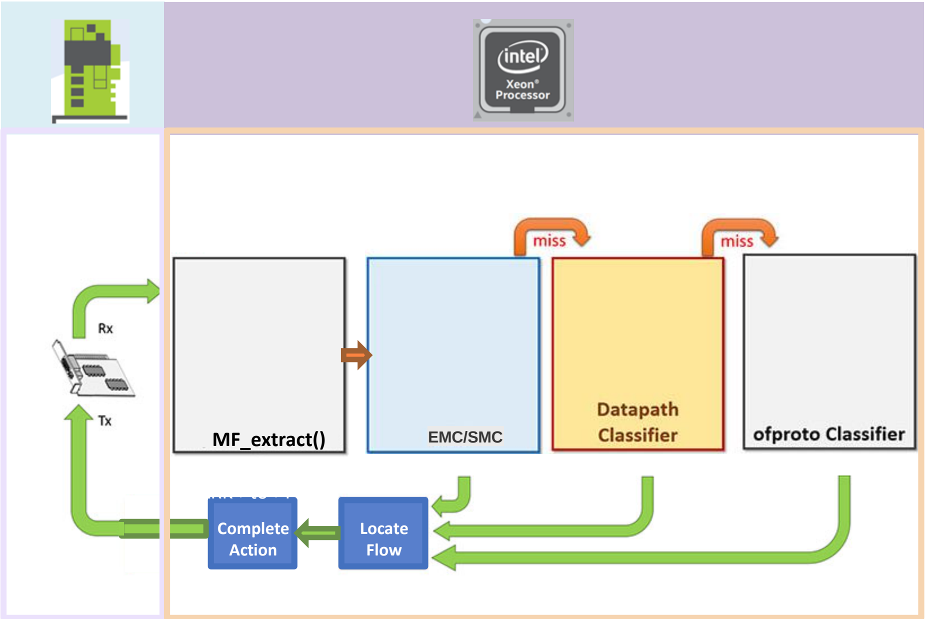
OVS: *hw-offload*

- Experimental feature, utilizing rte_flow, was first available in OVS 2.10 / Aug 2018*

```
ovs-vsctl set Open_vSwitch . other_config:hw-offload=true
```

- Only supports *partial offloading*
- Allows OVS to skip costly OPs
  - MF Extraction
  - EMC Lookup
  - DPCLS Lookup
- Protocols supported:

```
L2: Ethernet, VLAN - L3: IPv4, IPv6 - L4: TCP, UDP, SCTP, ICMP
```



(*) https://patchwork.ozlabs.org/cover/916146/

MF_Extract: Mini-Flow Extraction, EMC: Exact Match Cache, SMC: Signature Match Cache, ofproto: Open Flow Protocol

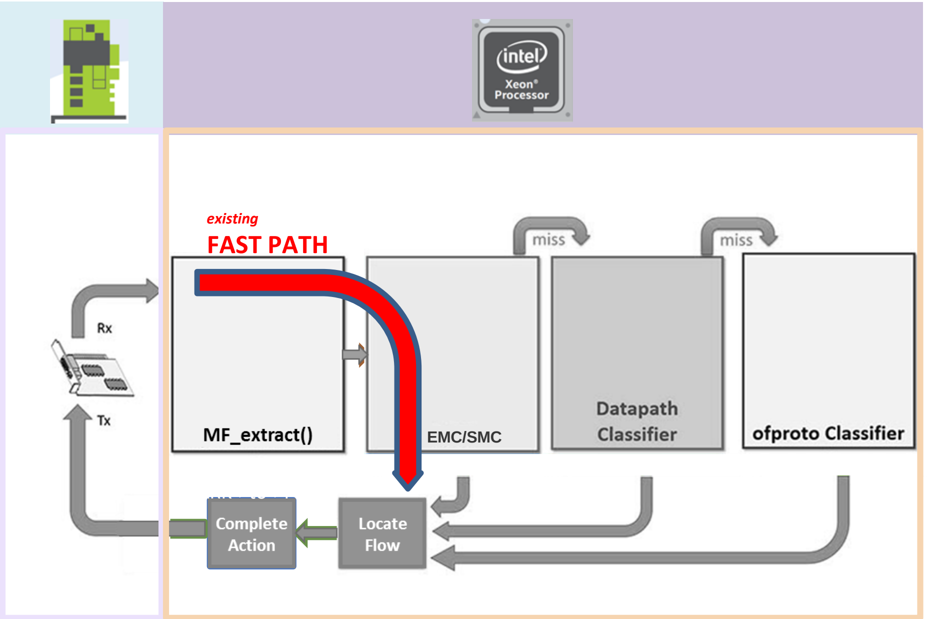
OVS: *hw-offload*

- Experimental feature, utilizing rte_flow, was first available in OVS 2.10 / Aug 2018*

```
ovs-vsctl set Open_vSwitch . other_config:hw-offload=true
```

- Only supports *partial offloading*
- Allows OVS to skip costly OPs
  - MF Extraction
  - EMC Lookup
  - DPCLS Lookup
- Protocols supported:

```
L2: Ethernet, VLAN - L3: IPv4, IPv6 - L4: TCP, UDP, SCTP, ICMP
```



(*) https://patchwork.ozlabs.org/cover/916146/

MF_Extract: Mini-Flow Extraction, EMC: Exact Match Cache, SMC: Signature Match Cache, ofproto: Open Flow Protocol

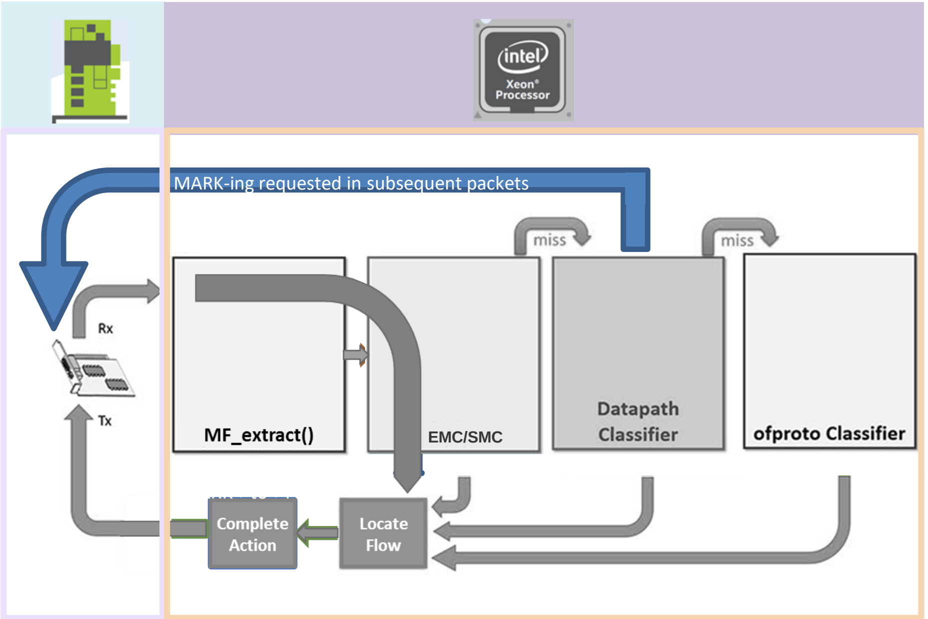
OVS: *hw-offload*

- Experimental feature, utilizing rte_flow, was first available in OVS 2.10 / Aug 2018*

```
ovs-vsctl set Open_vSwitch . other_config:hw-offload=true
```

- Only supports *partial offloading*
- Allows OVS to skip costly OPs
  - MF Extraction
  - EMC Lookup
  - DPCLS Lookup
- Protocols supported:

```
L2: Ethernet, VLAN - L3: IPv4, IPv6 - L4: TCP, UDP, SCTP, ICMP
```



[(*) https://patchwork.ozlabs.org/cover/916146/](https://patchwork.ozlabs.org/cover/916146/)

MF_Extract: Mini-Flow Extraction, EMC: Exact Match Cache, SMC: Signature Match Cache, ofproto: Open Flow Protocol

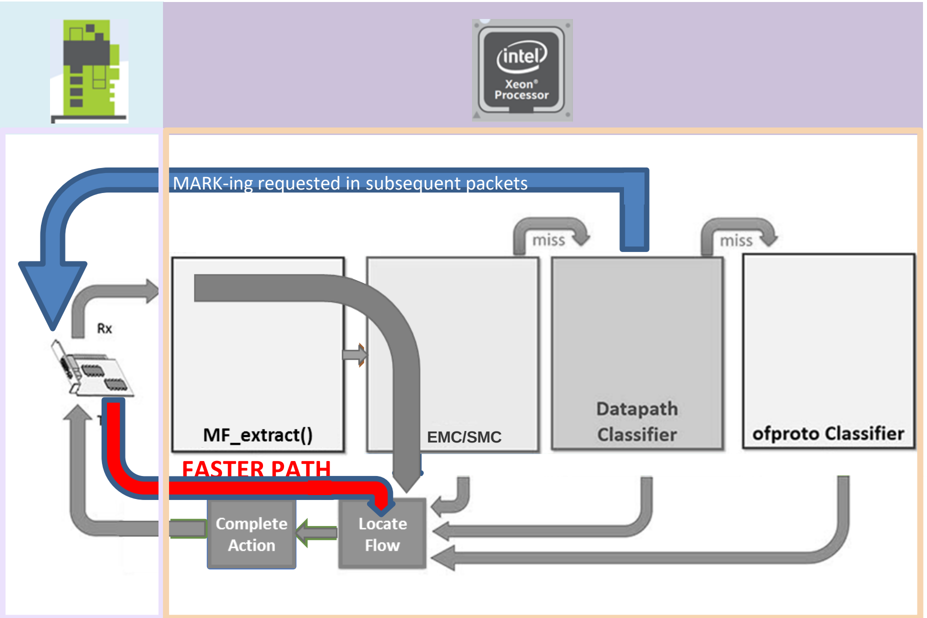
OVS: *hw-offload*

- Experimental feature, utilizing rte_flow, was first available in OVS 2.10 / Aug 2018*

```
ovs-vsctl set Open_vSwitch . other_config:hw-offload=true
```

- Only supports *partial offloading*
- Allows OVS to skip costly OPs
  - MF Extraction
  - EMC Lookup
  - DPCLS Lookup
- Protocols supported:

```
L2: Ethernet, VLAN - L3: IPv4, IPv6 - L4: TCP, UDP, SCTP, ICMP
```



(*) https://patchwork.ozlabs.org/cover/916146/

MF_Extract: Mini-Flow Extraction, EMC: Exact Match Cache, SMC: Signature Match Cache, ofproto: Open Flow Protocol

Intel ® Ethernet: *rte_flow & OVS hw-offload*

- Intel® Ethernet 700 Series Network Adapters (10/25/40GbE)
  - *rte_flow* driver in i40e PMD since flow API's inception (2017)
  - Utilizes Intel® Ethernet Flow Director feature in the controller
  - QUEUE, PASSTHRU, DROP, FLAG, MARK actions were available
  - Allows for up to 8K rules in device memory
- **NEW** in DPDK 19.08
  - Support for MARK + RSS actions added to rte_flow
- **NEW** in DPDK 19.11
  - Flow Director support for i40e vector RX path (SSE4.2 and AVX2)
- OVS hw-offload works with Intel® Ethernet 700 Series Controllers
  - Try in dpdk-next branch (now)
  - Out-of-box with OVS 2.13 (when released in 03/2020)
  - See application note https://bit.ly/2MbpJoA for more details.


X722-DA2
X722-DA4

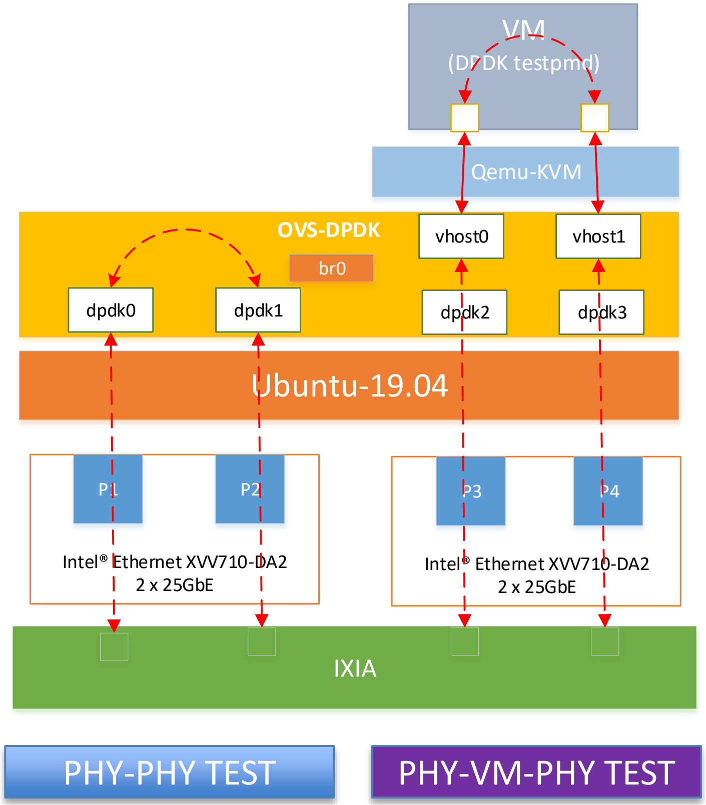

XXV710-DA1
XXV710-DA2


X710-DA2
X710-DA4FHBLK


XL710-QDA1
XL710-QDA2


X710-T4

OVS: *hw-offload Test Setup*



Hardware

Platform	Supermicro* X11DPH-T
CPU	**Intel(R) Xeon(R) Platinum 8180 CPU @ 2.50GHz**
MEMORY	Micron , DDR4 2666MHz, 16GB per Channel, 12 Channels, TOTAL: 192 GB
NIC	Intel® Ethernet XVV710-DA2 Adapter
BIOS	American Megatrends Inc., version 3.1 dated 05/22/2019

Software

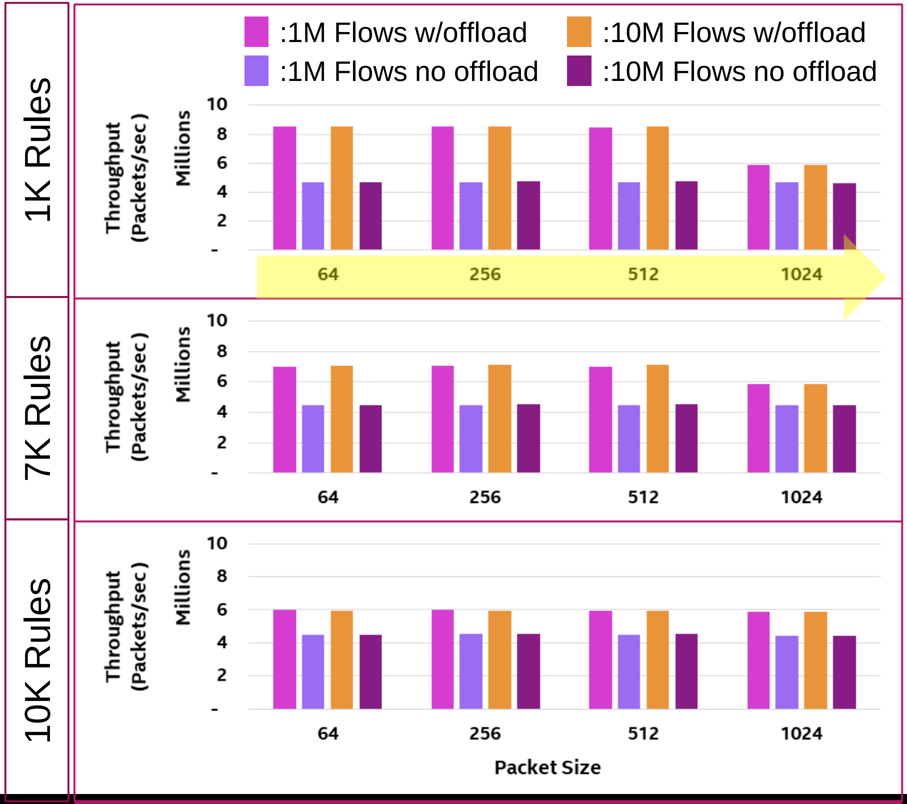
Host OS	Ubuntu* 19.04
Host Kernel version	5.0.0-23-generic
BIOS settings	P-state Disabled, C-States Disabled
SW Version	DPDK 19.11-rc1, OVS 2.11.1 (dpdk-latest)
IXIA TEST	RFC 2544 0.01% PACKET LOSS

Configurations

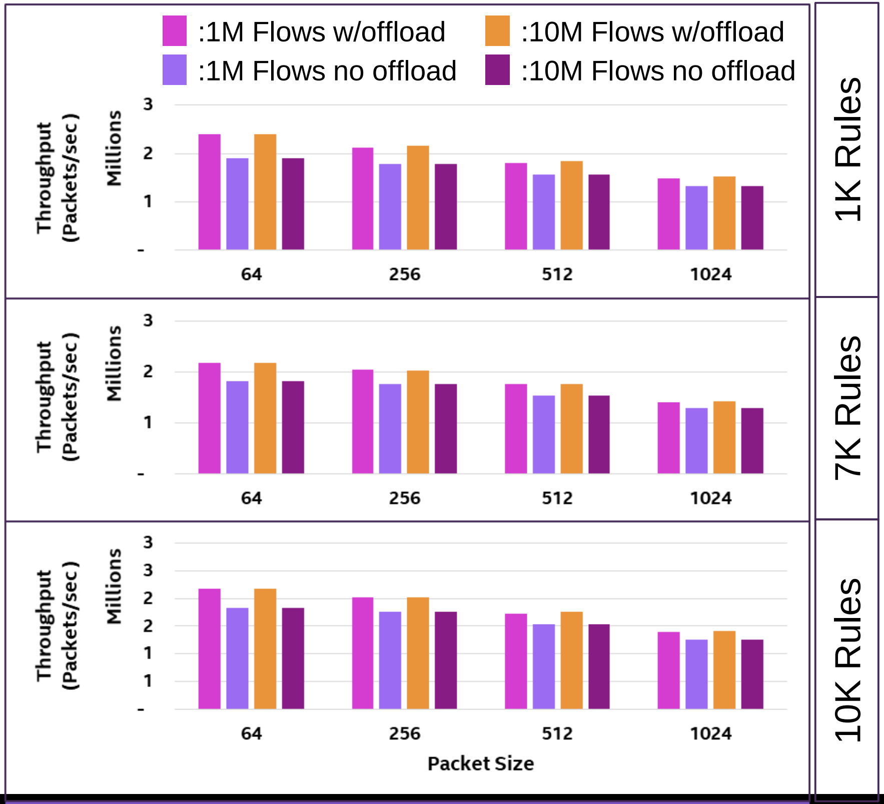
Flows (Millions)	1, 10
Rules (Thousands)	1, 7, 10
Packet Size (Bytes)	64, 256, 512, 1024
Protocols Matched	IP, UDP, TCP
CPU Cores Used	1

OVS: *hw-offload Throughput Experiment Results*

OVS: *hw-offload Throughput Experiment Results*



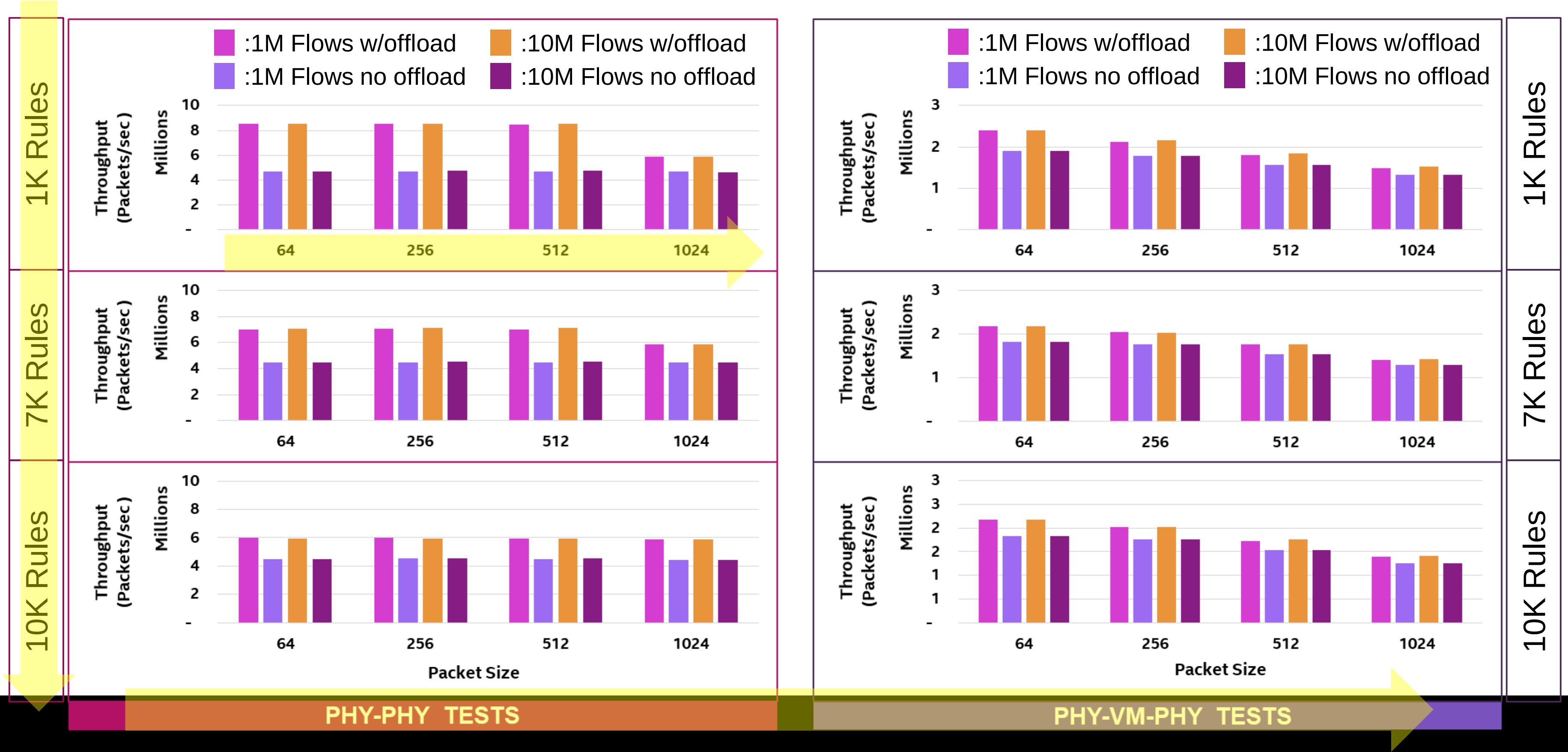
PHY-PHY TESTS



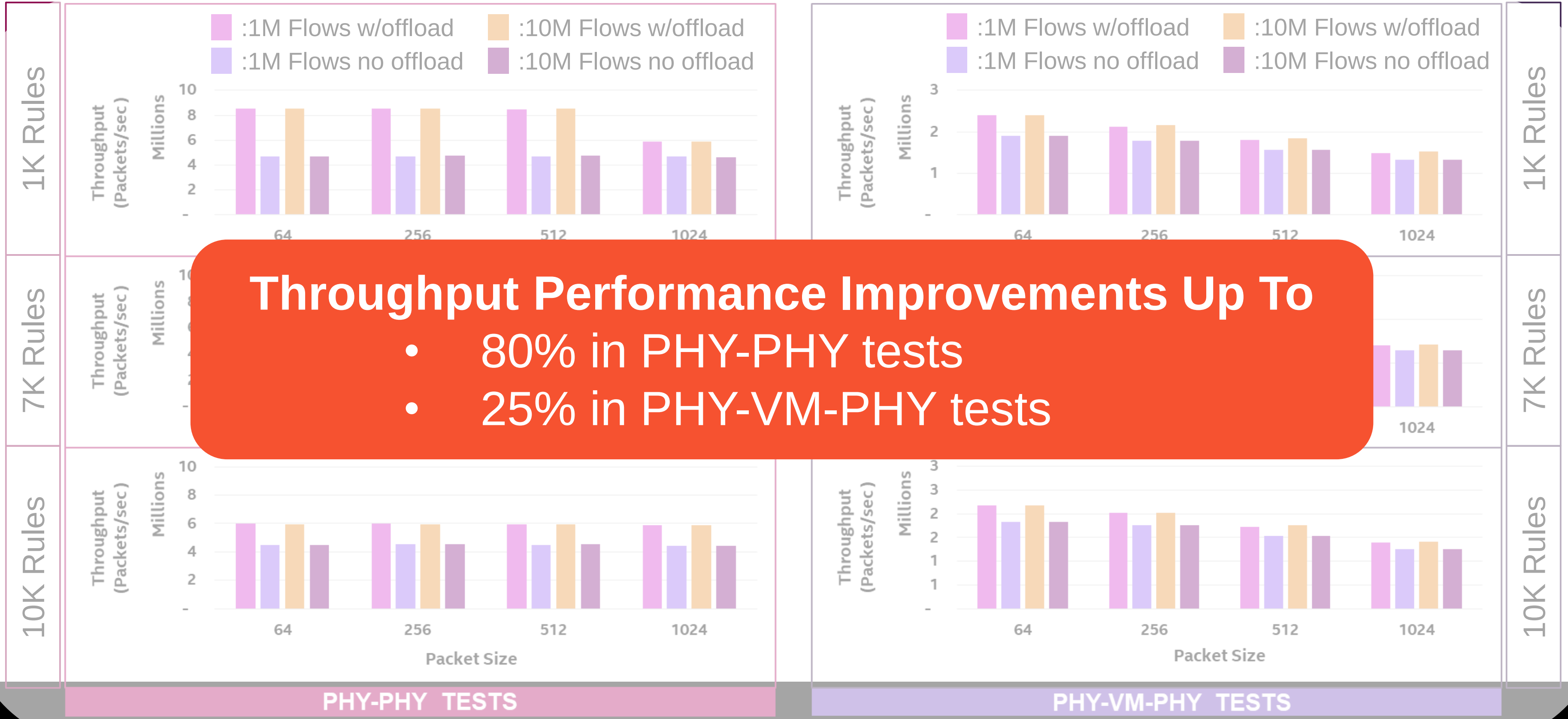
PHY-VM-PHY TESTS

OVS: *hw-offload Throughput Experiment Results*

OVS: *hw-offload Throughput Experiment Results*



OVS: *hw-offload Throughput Experiment Results*

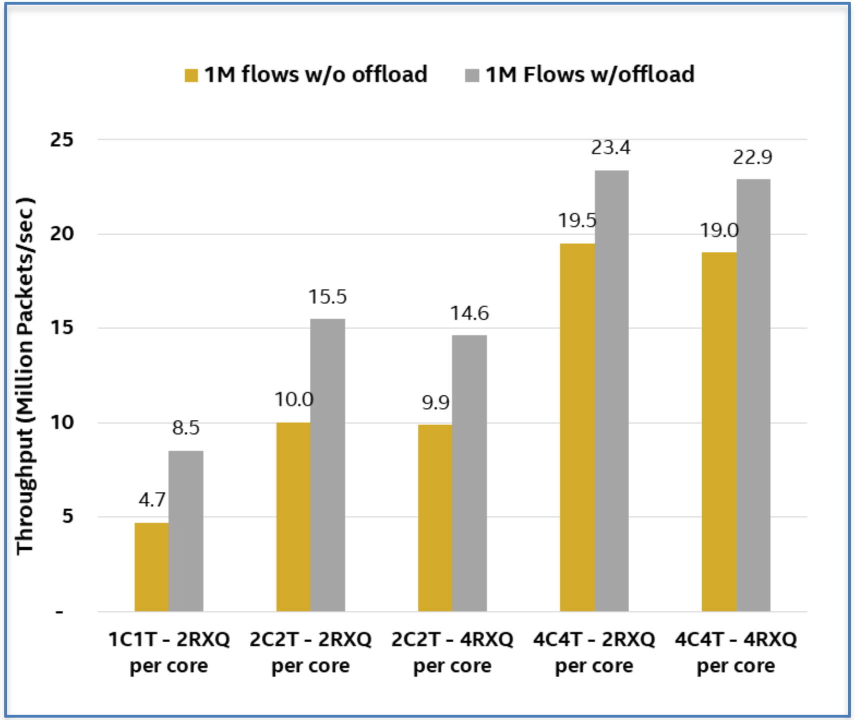


Throughput Performance Improvements Up To

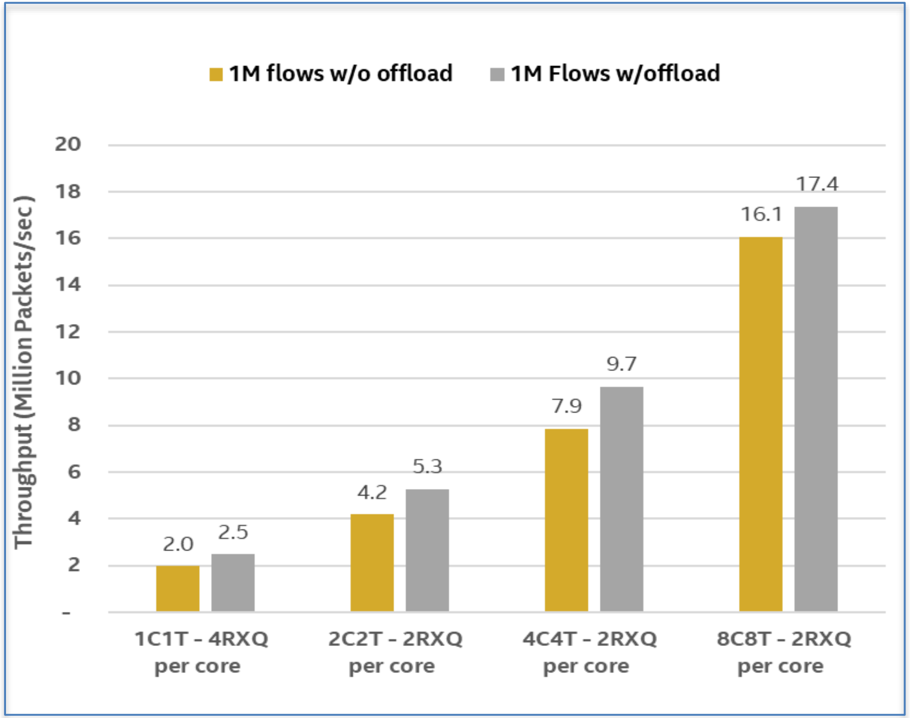
- 80% in PHY-PHY tests
- 25% in PHY-VM-PHY tests

OVS: *hw-offload Core Scaling Experiment Results*

1K Rules (64B)



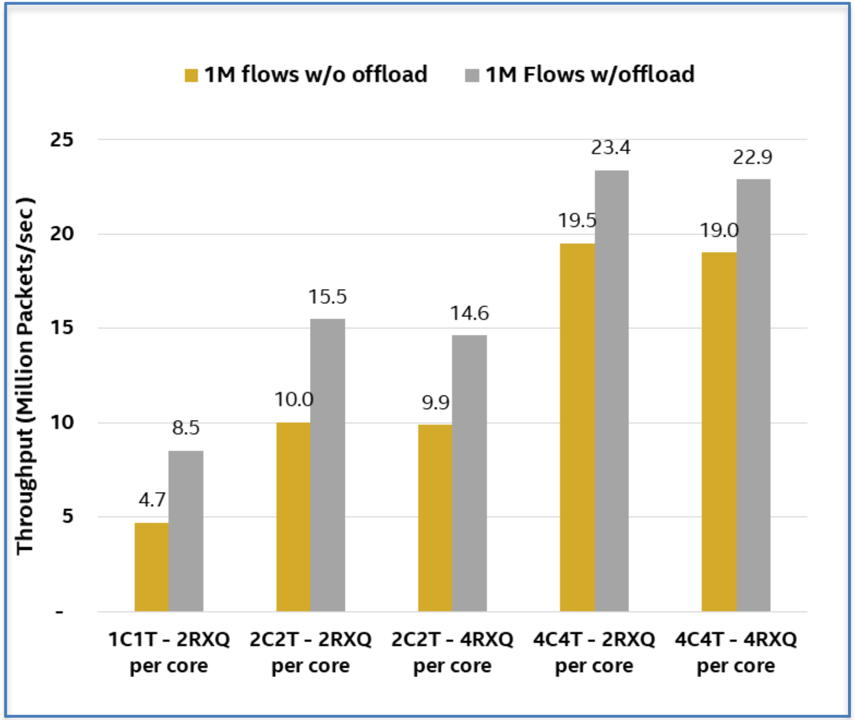
PHY-PHY TEST



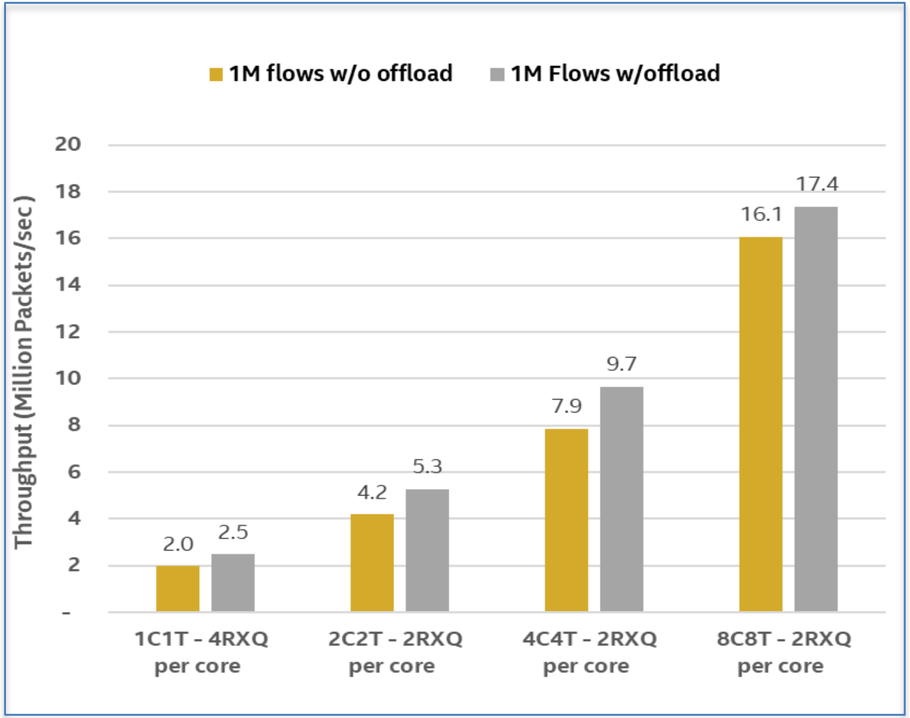
PHY-VM-PHY TEST

OVS: *hw-offload Core Scaling Experiment Results*

1K Rules (64B)



PHY-PHY TEST



PHY-VM-PHY TEST

Summary

- OVS DPDK hw-offload is
  - available on physical ports, now including Intel Ethernet 700 Series Devices
  - not supported on vhost-backend and virtio
  - able to offload rules matching on Ethernet, IP, TCP, and UDP protocols
  - Intel Ethernet 700 Series Devices can hold up to 8K rules in device memory

- Performance improvements are
  - due to cycles made available to flow access and action processing
  - consistent across packet sizes
  - proportional to the share of cycles for physical ports
  - diminishing as you use four or more cores per physical

The End

References:

DPDK Programmers Guide @ https://doc.dpdk.org/guides/prog_guide

Open vSwitch Hardware Acceleration Application Note @ https://bit.ly/2MbpJoA

HOWTO: Using OVS DPDK @ http://docs.openvswitch.org/en/latest/howto/dpdk

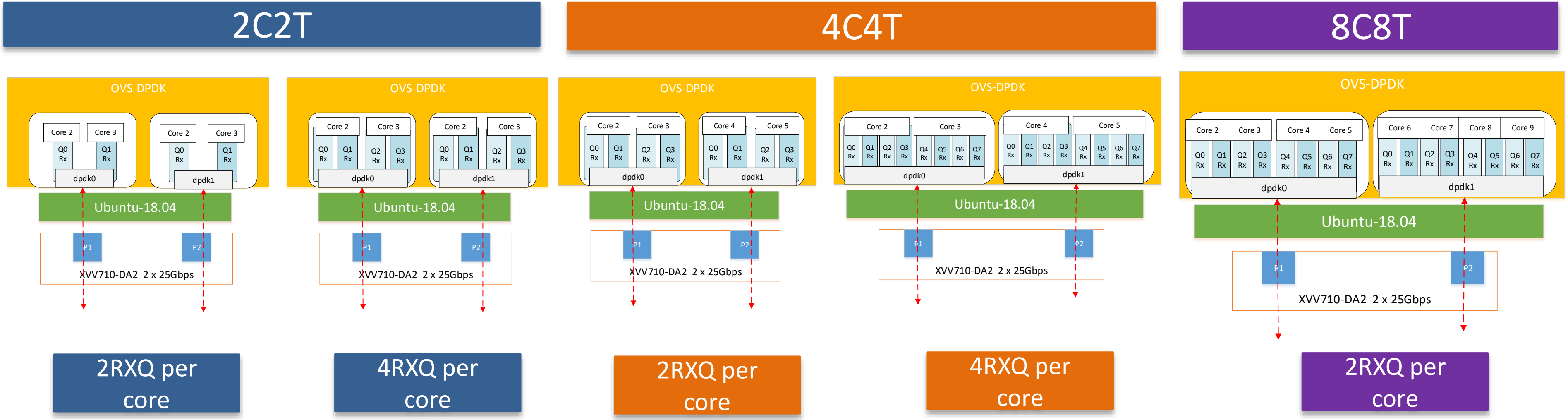
Acknowledgements:

Harry Van Haaren, Irene Liew, Chenmin Sun, Yipeng Wang, Charlie Tai, John McNamara, Edwin Verplanke

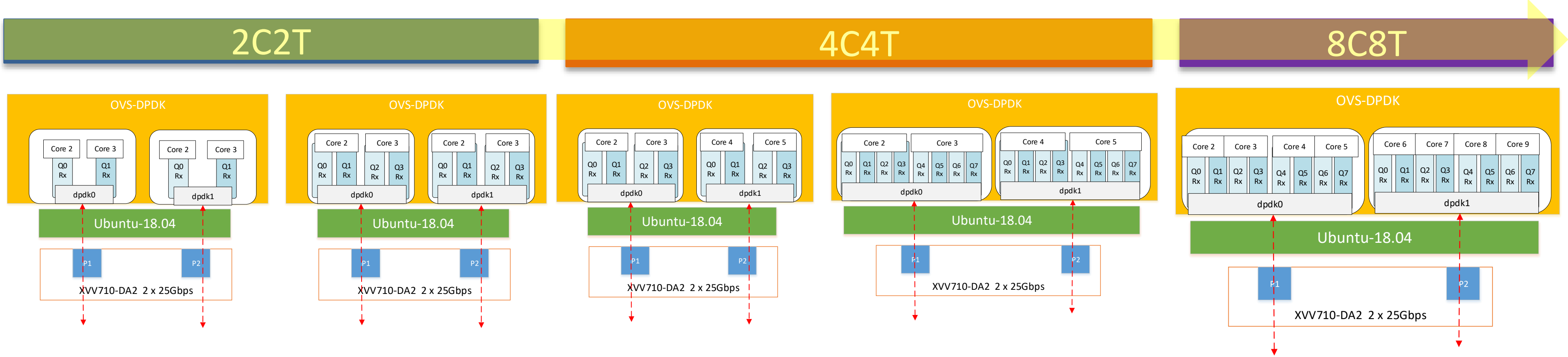
Contact us at:

npl@intel.com

OVS-DPDK Phy-Phy Core Scaling Test Configuration



OVS-DPDK Phy-Phy Core Scaling Test Configuration



2RXQ per core

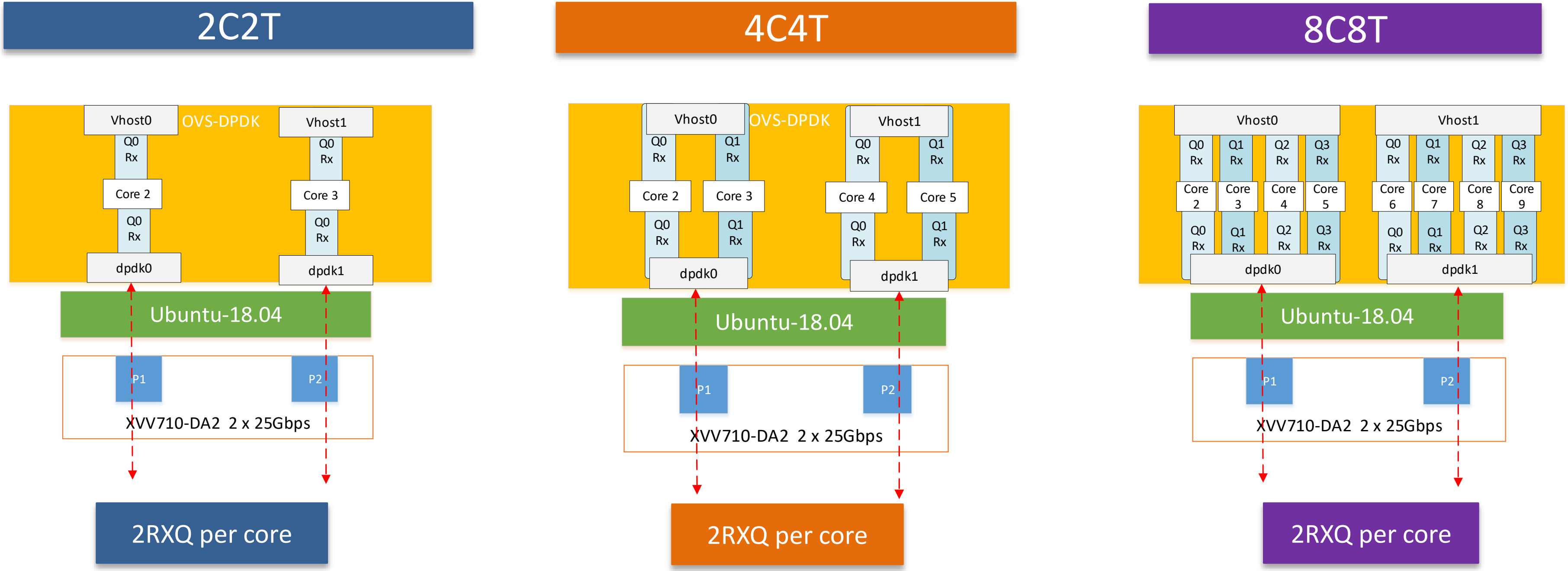
4RXQ per core

2RXQ per core

4RXQ per core

2RXQ per core

OVS-DPDK Phy-VM-Phy Core Scaling Test Configuration



OVS-DPDK Phy-VM-Phy Core Scaling Test Configuration