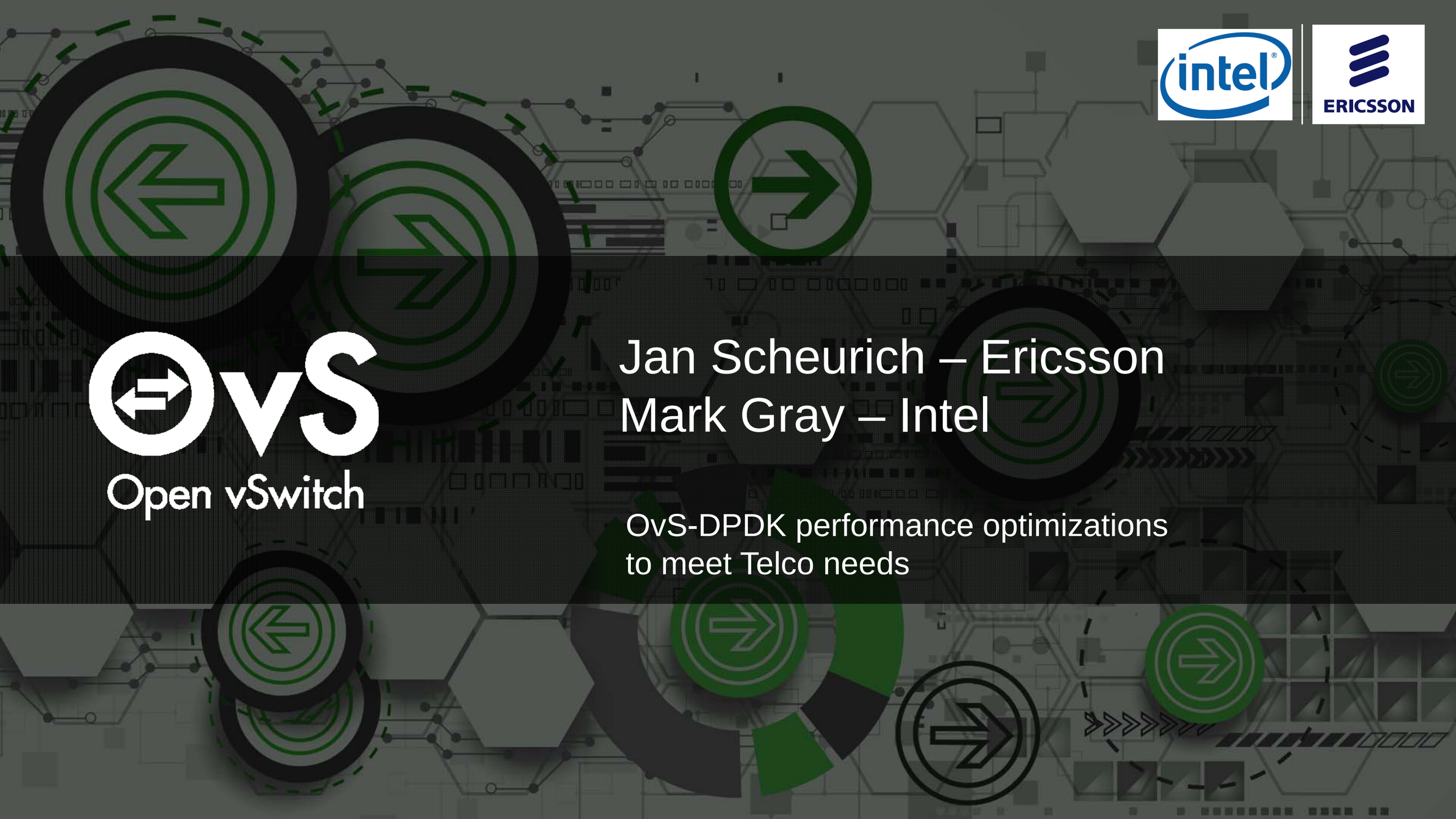


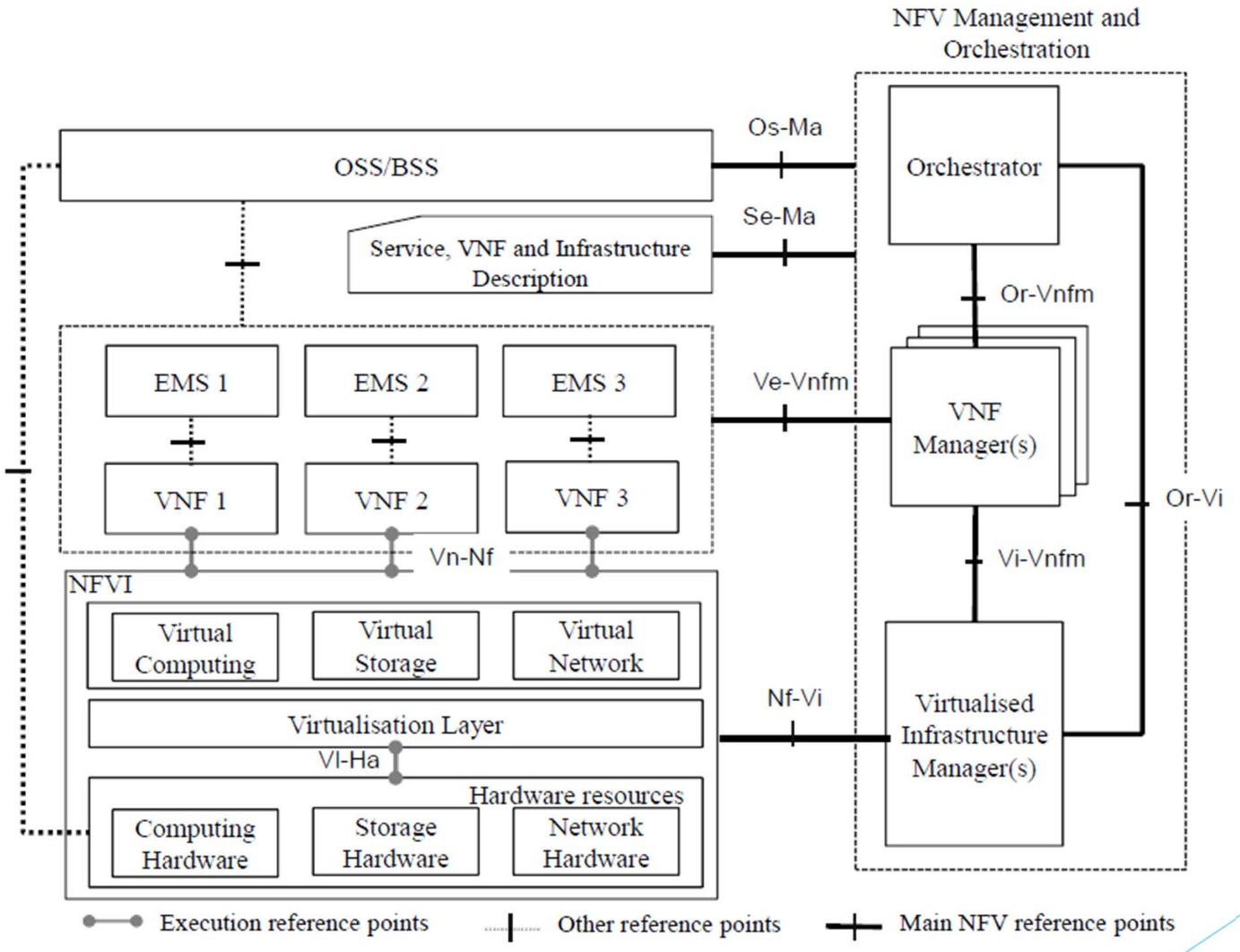
Jan Scheurich – Ericsson
Mark Gray – Intel

OvS-DPDK performance optimizations to meet Telco needs

Introduction

- OVS-DPDK in complex NFV environments
- What determines performance in OVS-DPDK?
- OVS 2.5 performance baseline in L3-VPN use case
- Find and address performance bottlenecks
- Achieved improvements in OVS 2.6 and beyond
- Potential future work

What is NFV?


NFV Management and Orchestration
OSS/BSS
Os-Ma
Orchestrator
Se-Ma
Service, VNF and Infrastructure Description
Or-Vnfm
EMS 1
EMS 2
EMS 3
Ve-Vnfm
VNF Manager(s)
VNF 1
VNF 2
VNF 3
Or-Vi
Vn-Nf
Vi-Vnfm
NFVI
Virtual Computing
Virtual Storage
Virtual Network
Nf-Vi
Virtualised Infrastructure Manager(s)
Virtualisation Layer
Vl-Ha
Hardware resources
Computing Hardware
Storage Hardware
Network Hardware
Execution reference points
Other reference points
Main NFV reference points

Virtual Network Functions

Firewall

Load Balancer

Deep Packet Inspection

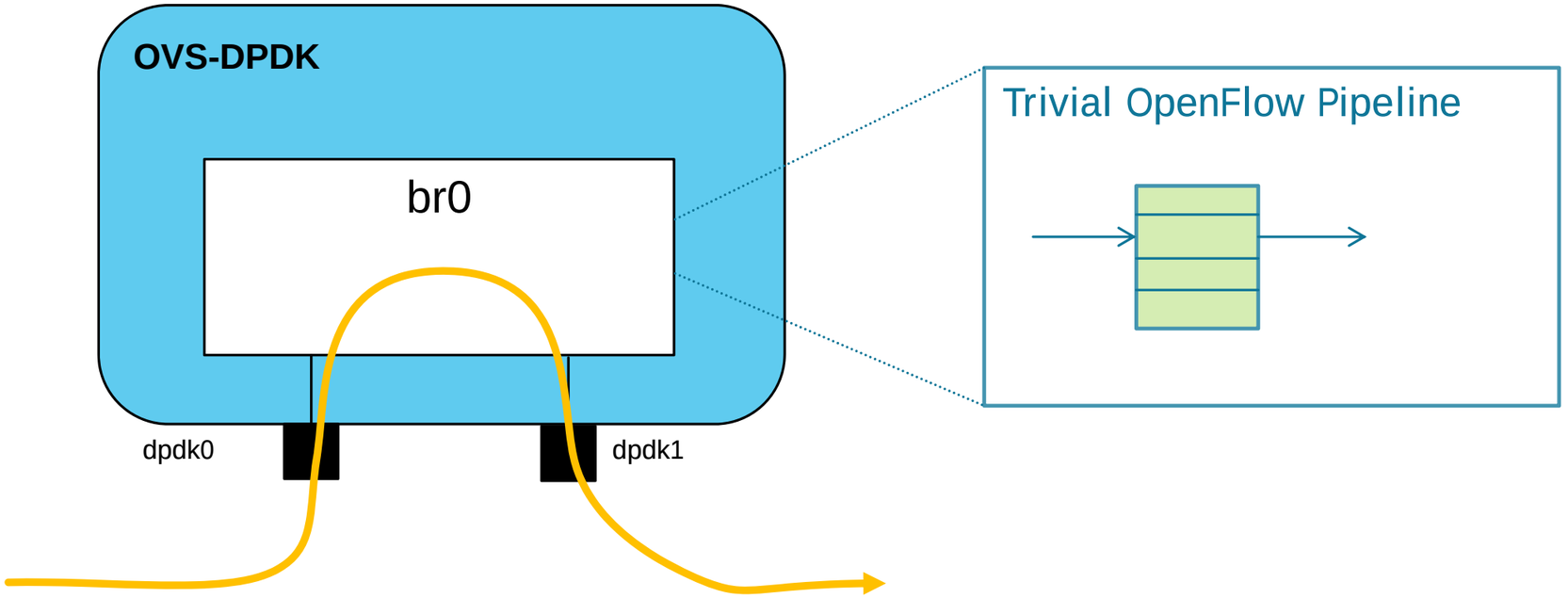
Content Filter

Carrier Grade Network Address Translation

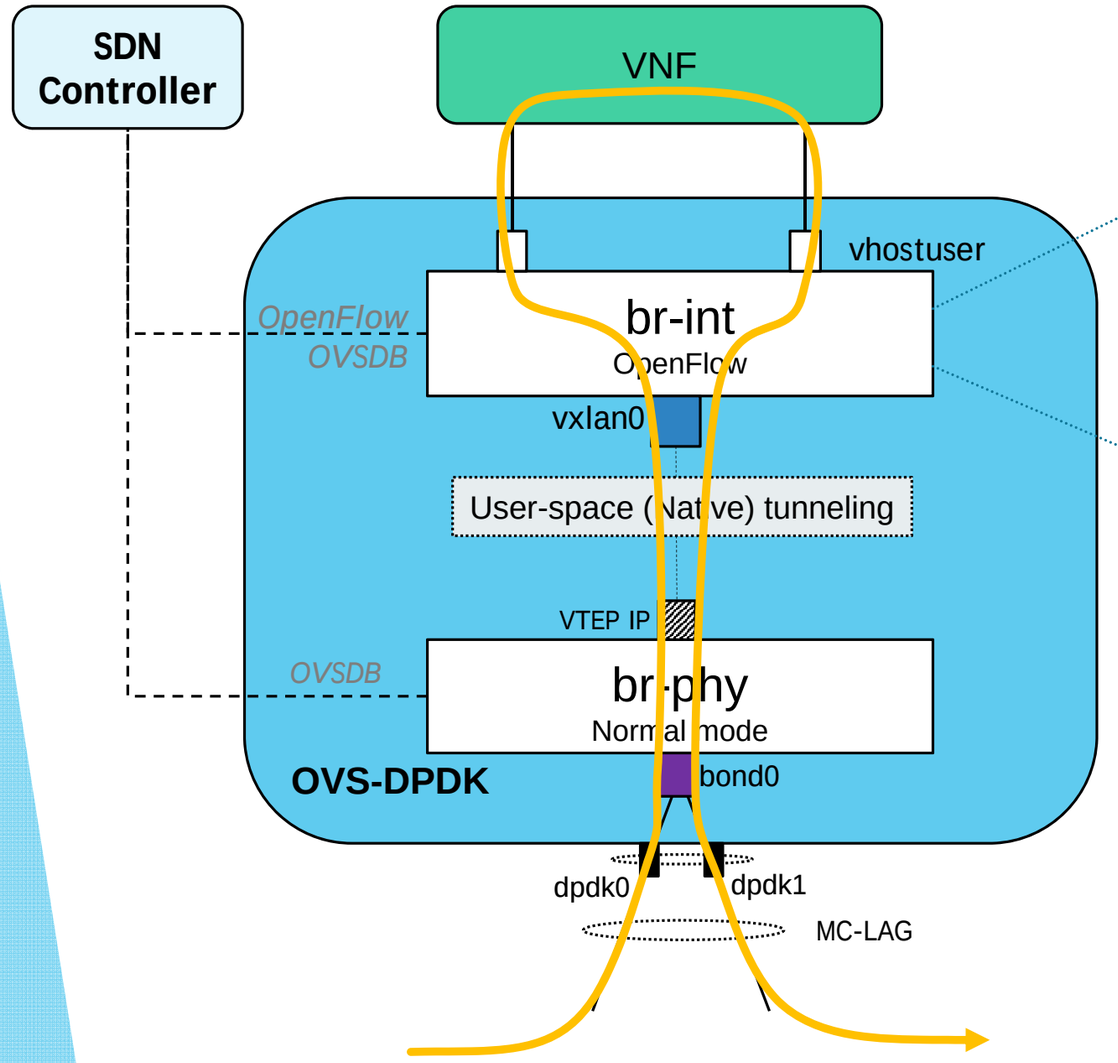
Evolved Packet Gateway

OvS Open vSwitch - DPDK

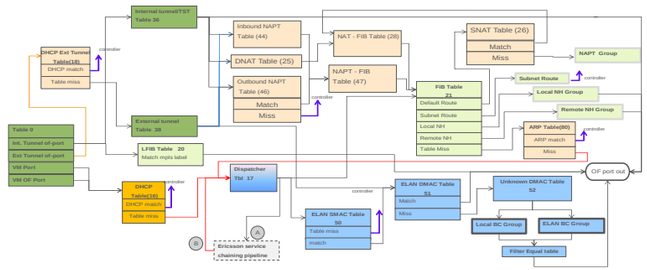
Typical OVS Benchmark Setup



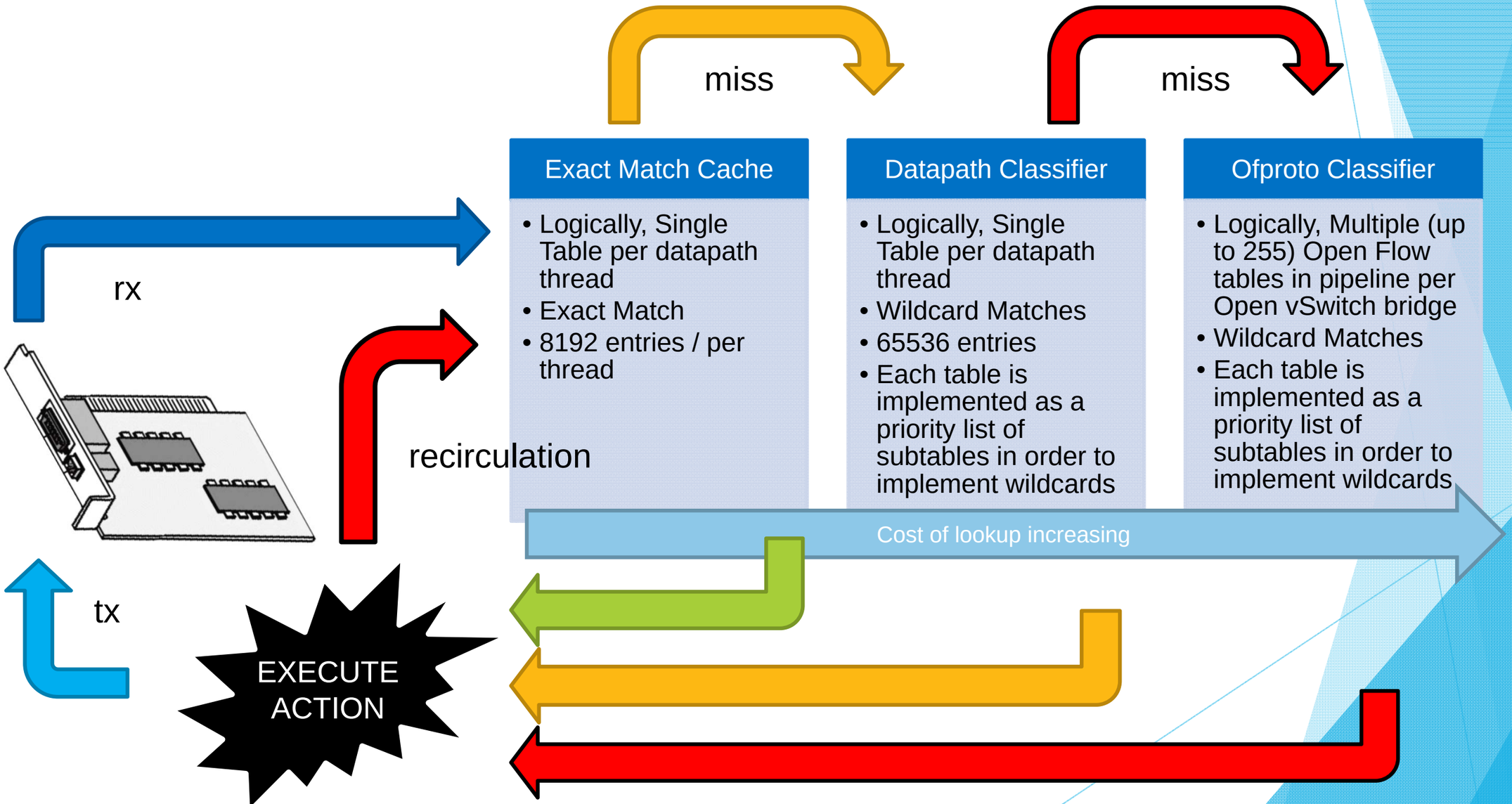
Typical OVS Configuration for NFV



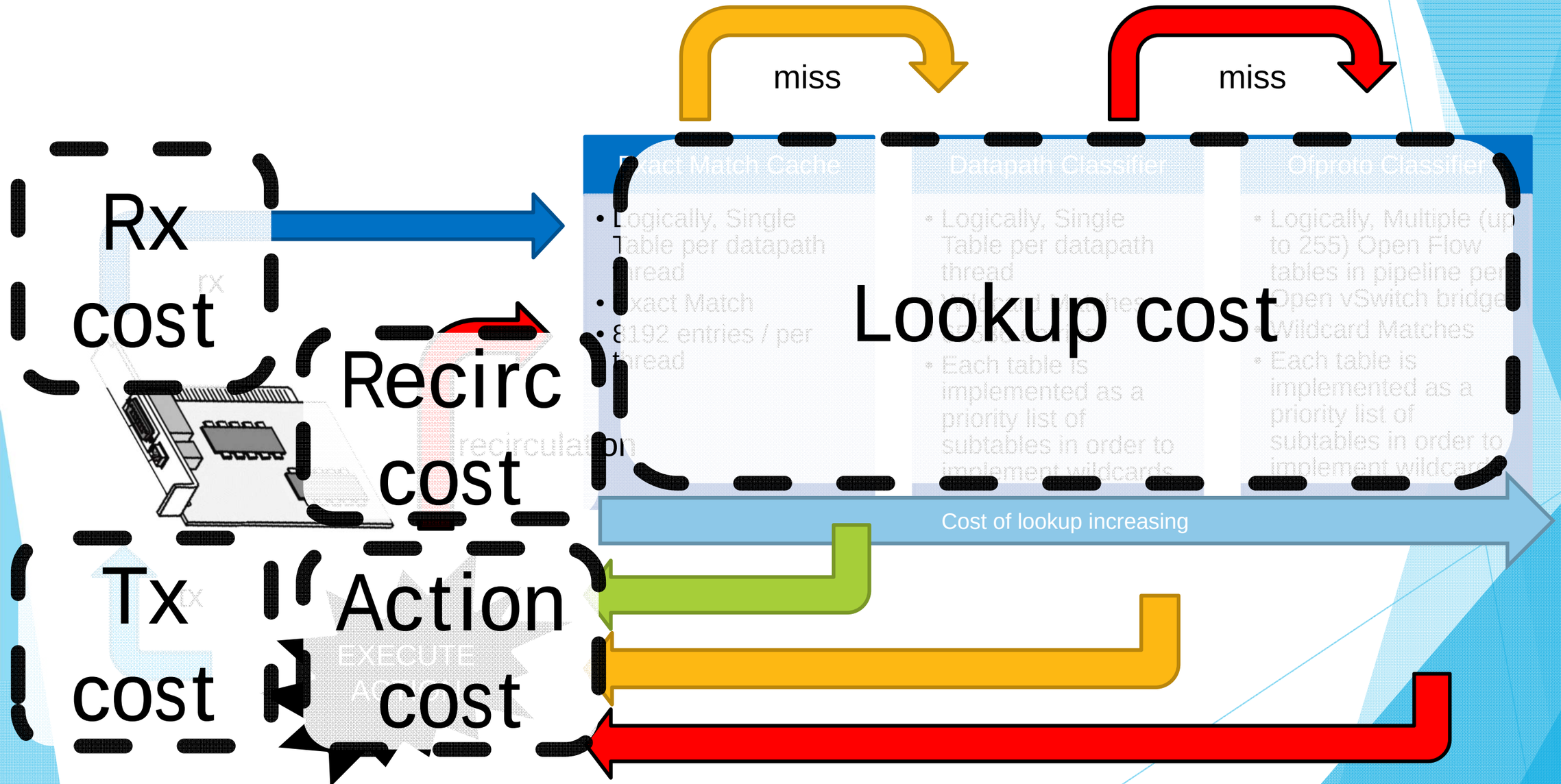
Complex OpenFlow pipelines



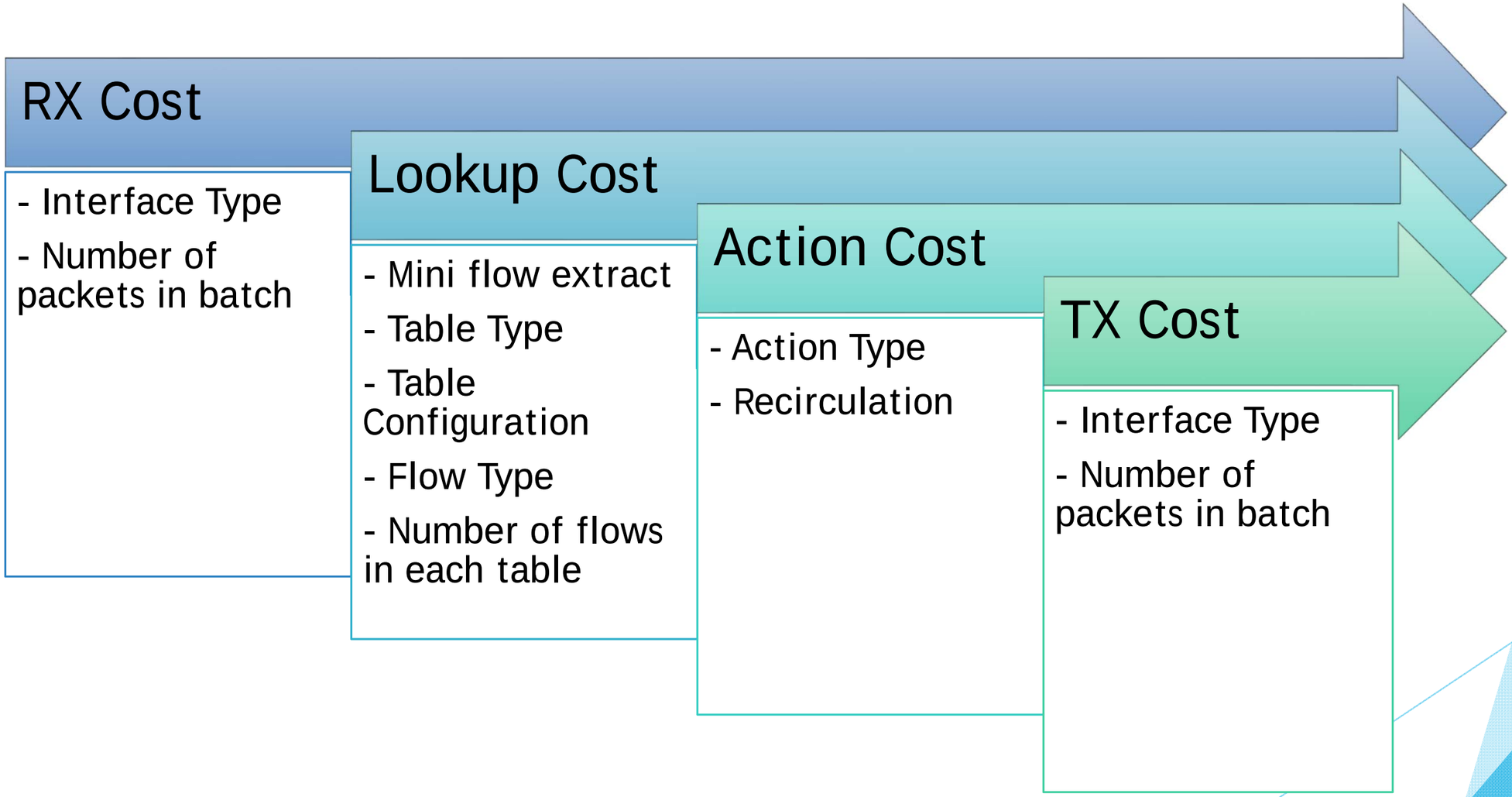
What affects OVS-DPDK performance?



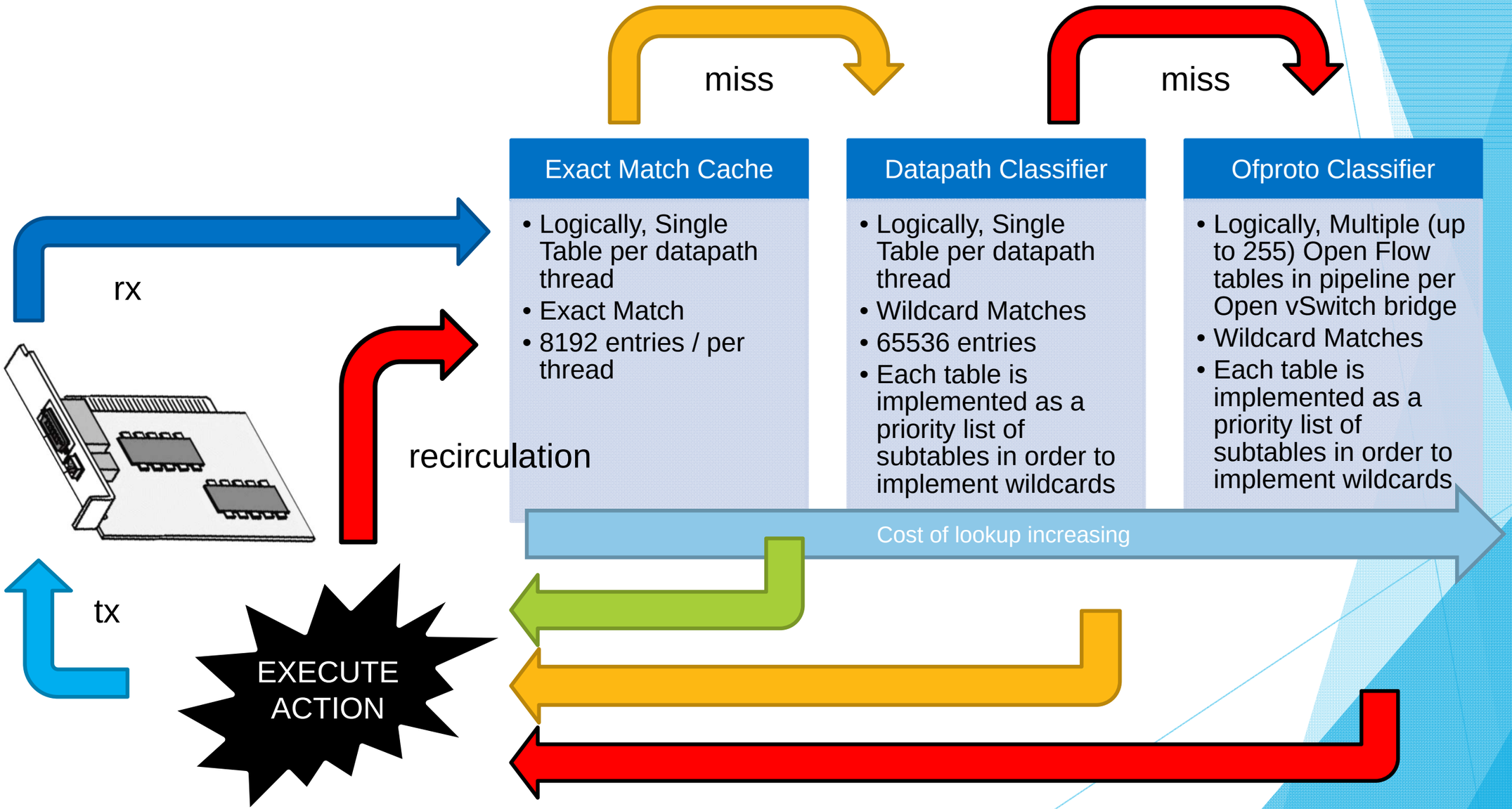
What affects OVS-DPDK performance?



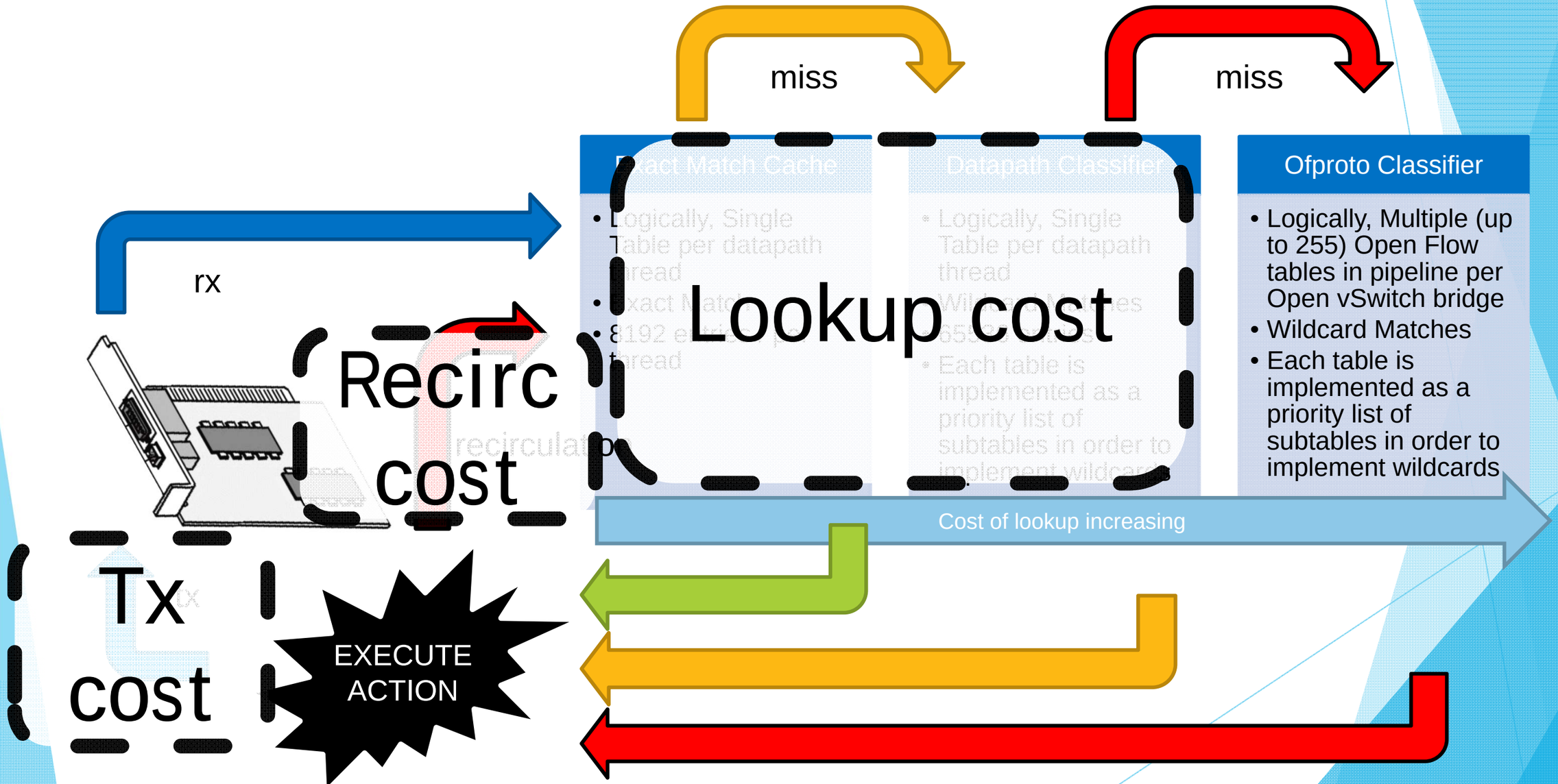
What affects OVS-DPDK performance?



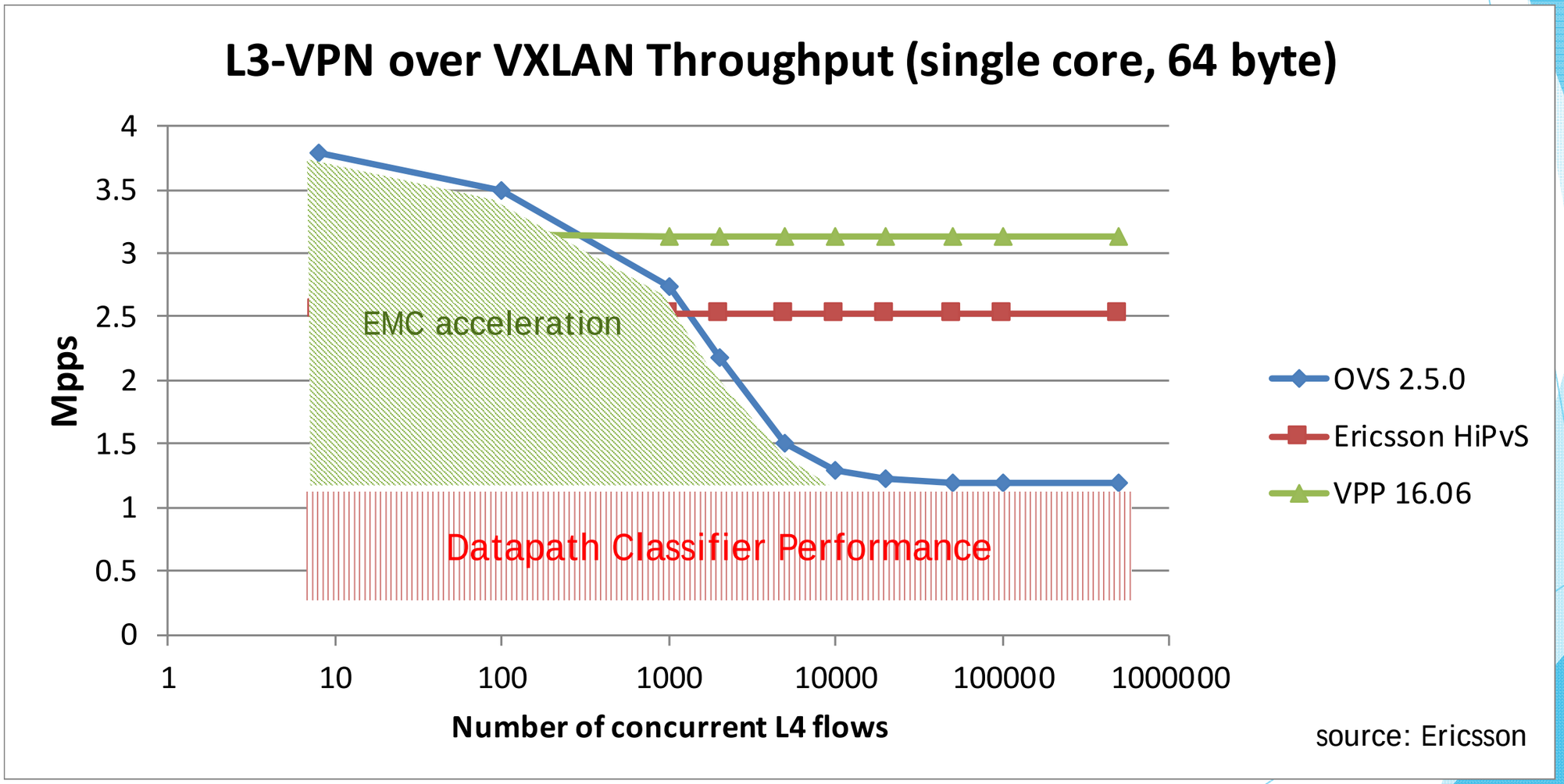
What affects OVS-DPDK performance?



What this work focuses on:



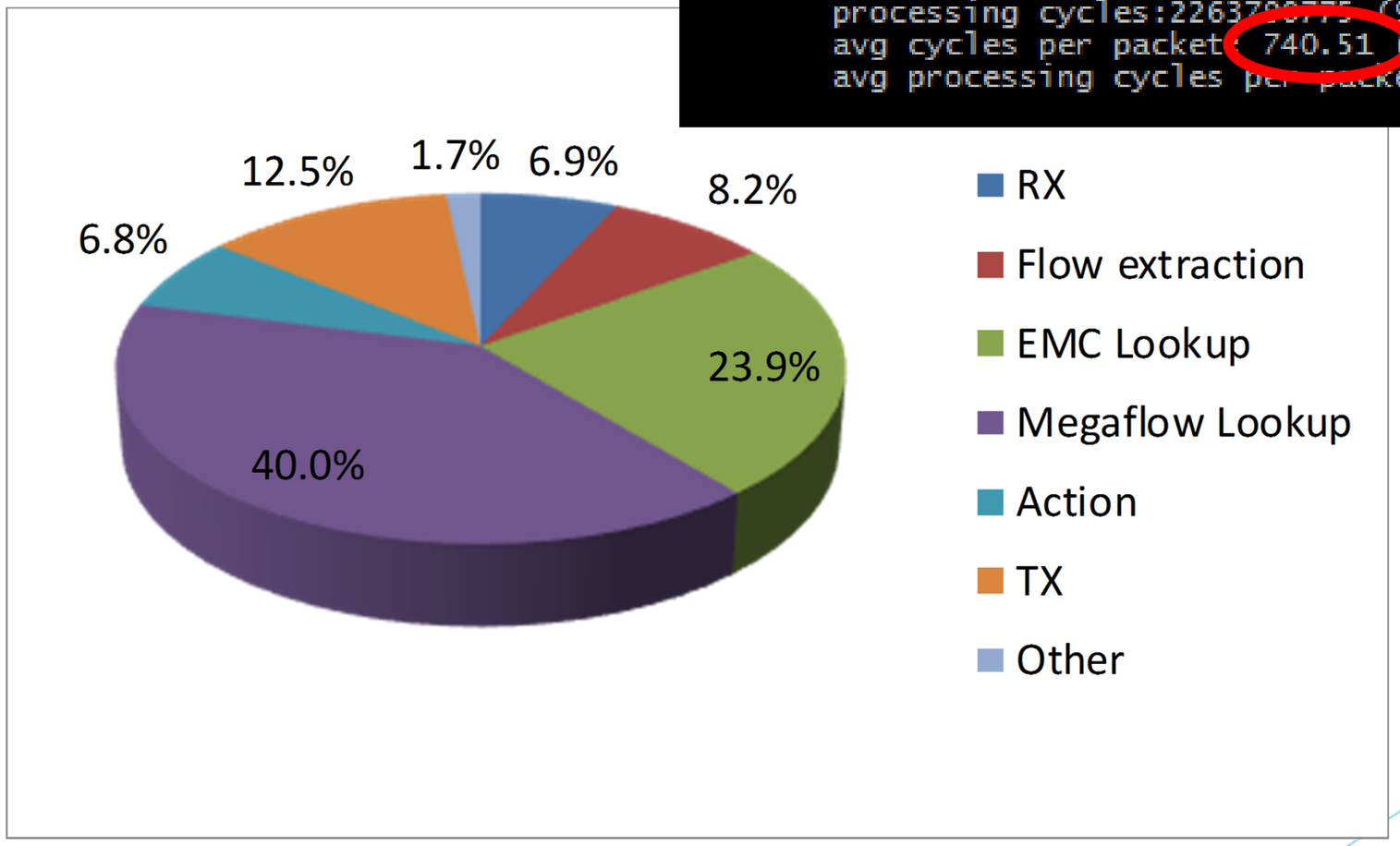
Ericsson Benchmark: Performance Baseline: OVS 2.5.0



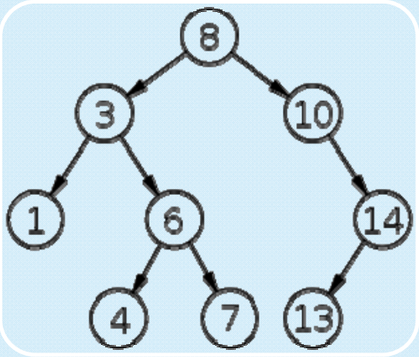
CPU: Single socket, Xeon CPU E5-2658 v2 @ 2.40GHz, 10 cores + HT, 640K L1, 2560K L2, 25MB L3 cache
NIC: Intel 82599, 2 x 10Gigabit/s, Memory: 4 banks of 16GB DDR3 1600 MHz

Cost Breakdown of L3-VPN in OVS 2.5
(4000 L4 flows)

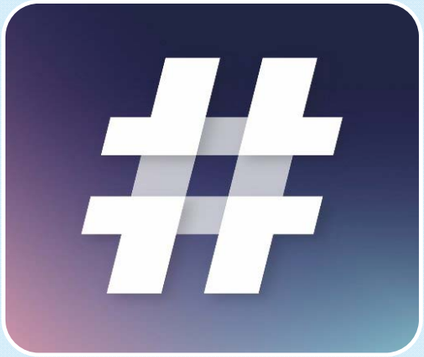
```
pmd thread numa_id 0 core_id 1:
        emc hits:1512270
        megaflow hits:1732461
        miss:0
        lost:0
        polling cycles:138949317 (5.78%)
        processing cycles:2263790775 (94.22%)
        avg cycles per packet: 740.51 (2402740092/3244731)
        avg processing cycles per packet: 697.68 (2263790775/3244731)
```



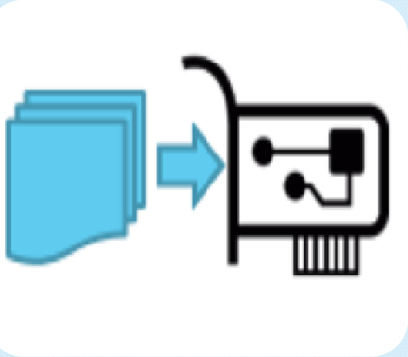
Optimization Activities (1/2)



Replace tuple space classifier with a trie based classifier



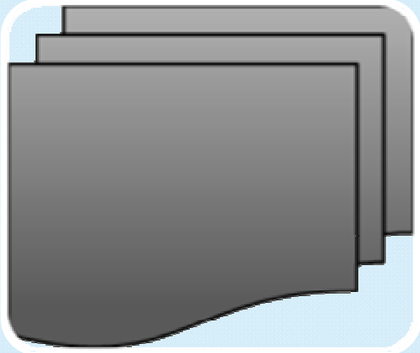
Faster crc32 hash function



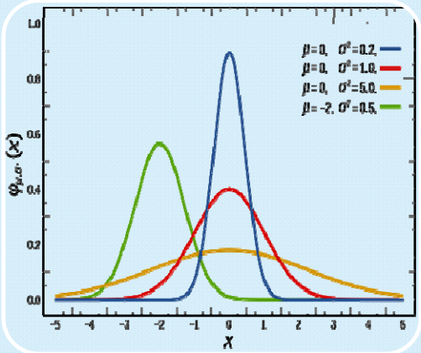
TX packet batching

Data structure alignment

Optimization Activities (2/2)

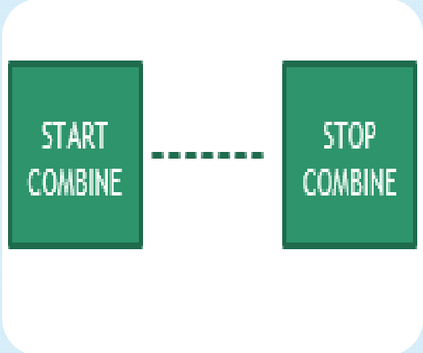


dpcls per in_port with sorted subtables



Probabilistic EMC insertion

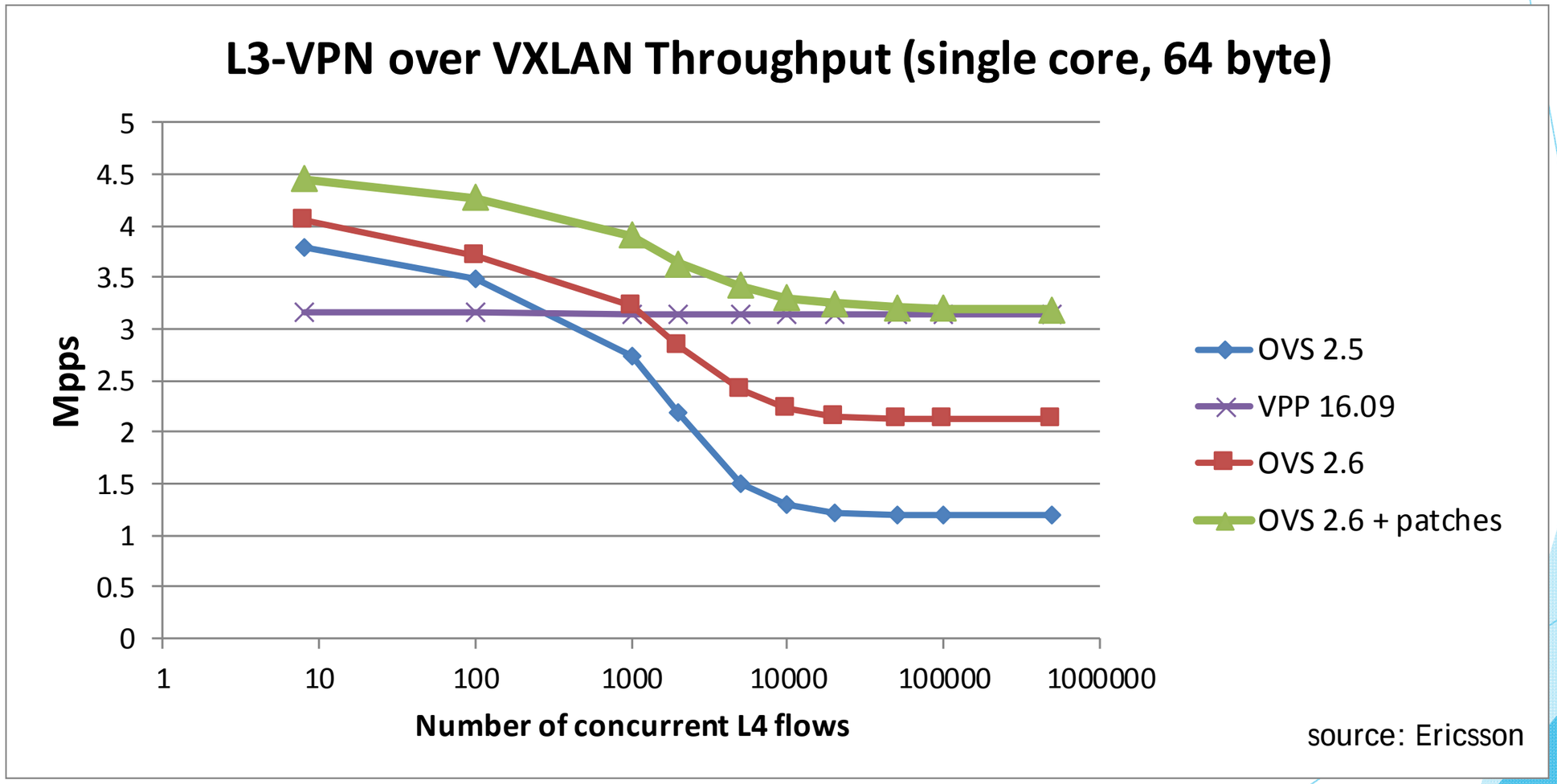
More meaningful PMD performance debug info



Combine actions for TX to tunnel to avoid recirculation

OVS 2.6

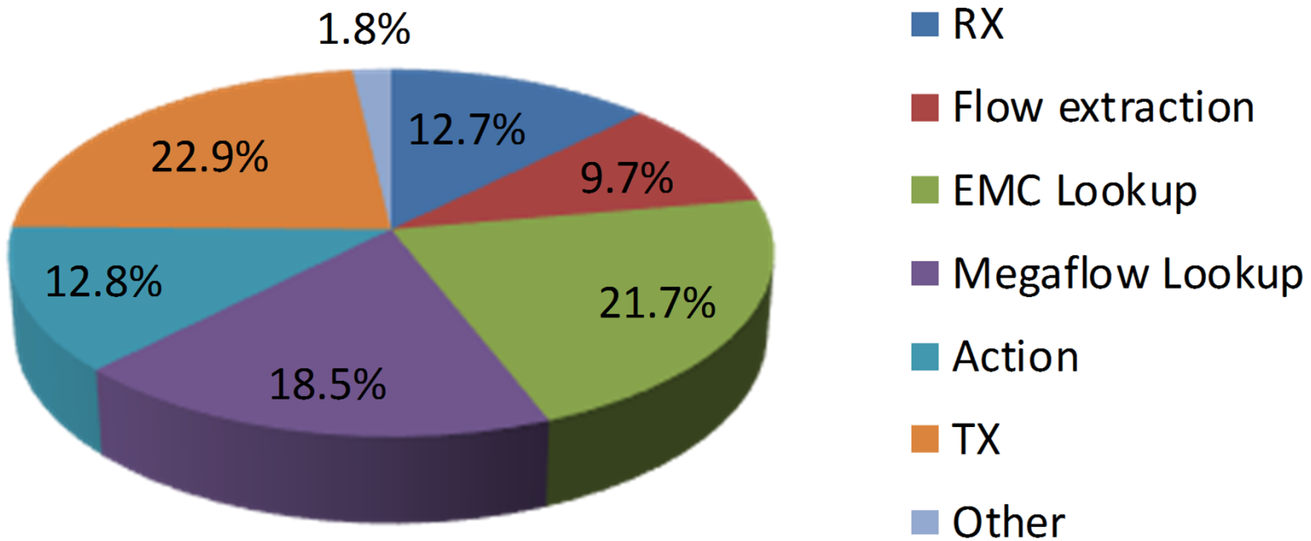
Ericsson Benchmark:
OVS Performance Improvements



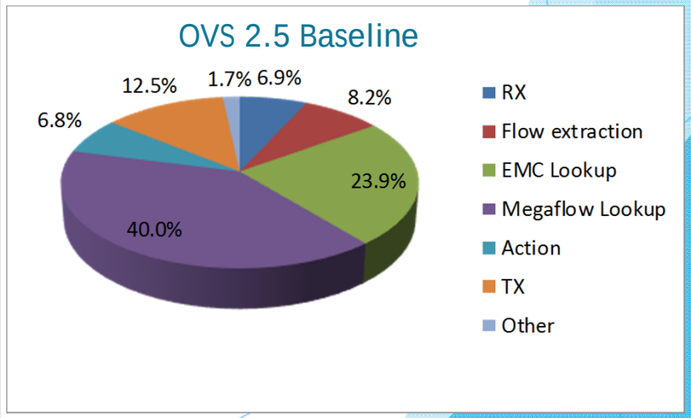
CPU: Single socket, Xeon CPU E5-2658 v2 @ 2.40GHz, 10 cores + HT, 640K L1, 2560K L2, 25MB L3 cache
NIC: Intel 82599, 2 x 10Gigabit/s, Memory: 4 banks of 16GB DDR3 1600 MHz

Cost Breakdown after Optimizations
(4000 L4 flows)

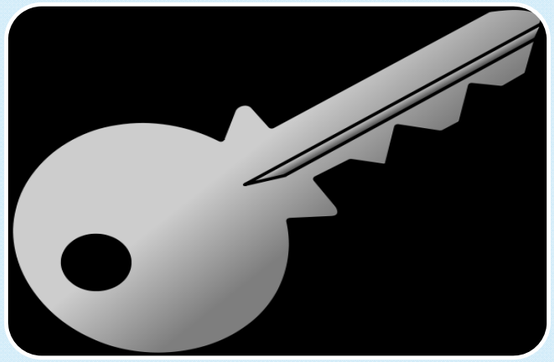
```
pmd thread numa_id 0 core_id 1:
        emc entries:7429 (90.69% full)
        emc hits:2949720
        megaflow hits:2158193
        avg. subtable lookups per hit:1.00
        miss:0
        lost:0
        idle cycles:0 (0.00%)
        processing cycles:2418234753 (100.00%)
        avg cycles per packet: 332.81 (2418234753/7266106)
        avg processing cycles per packet: 332.81 (2418234753/7266106)
```



source: `perf top`



Future Efforts



Lookup key on demand

Action cost reduction

Others?

Summary

- OVS-DPDK is being deployed as a virtual switch in complex NFV environments
- Exposes OVS to more complex configurations and traffic profiles than in traditional use cases
- Targeted optimization and redesign efforts have successfully improved the performance of OVS-DPDK for a typical NFV use case by a factor of 2.6
- Collaboration between teams with different experiences and viewpoints can yield great results!

Disclaimers

Intel does not control or audit third-party benchmark data or the web sites referenced in this document. You should visit the referenced web site and confirm whether referenced data are accurate.

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].**

Questions?

References

- DPCLS per in_port with sorted subtables
  commit 3453b4d62a98f1c276a89ad560d4212b752c7468
- Data structure alignment
  http://openvswitch.org/pipermail/dev/2016-October/080654.html
- Probabilistic EMC insertion
  http://openvswitch.org/pipermail/dev/2016-November/xxxxx/html
- PMD performance debug info
  http://openvswitch.org/pipermail/dev/2016-November/xxxxx/html
- TX Batching
  http://openvswitch.org/pipermail/dev/2016-November/xxxxx/html
- TX to tunnel ports without recirculation (combine actions)
  http://openvswitch.org/pipermail/dev/2016-November/xxxxx/html