Are You Insured Against Your Noisy Neighbor

Sunku Ranganath, Intel Corporation
Sridhar Rao, Spirent Communications

@SunkuRanganath, @ngignir

Legal Disclaimer

Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.

Intel processors of the same SKU may vary in frequency or power as a result of natural variability in the production process.

For more complete information about performance and benchmark results, visit www.intel.com/benchmarks.

The cost reduction scenarios described are intended to enable you to get a better understanding of how the purchase of a given Intel based product, combined with a number of situation-specific variables, might affect future costs and savings. Circumstances will vary and there may be unaccounted-for costs related to the use and deployment of a given product. Nothing in this document should be interpreted as either a promise of or contract for a given level of costs or cost reduction.

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.

Optimization Notice: Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice. Notice Revision #20110804.

No computer system can be absolutely secure.

Intel® Advanced Vector Extensions (Intel® AVX)* provides higher throughput to certain processor operations. Due to varying processor power characteristics, utilizing AVX instructions may cause a) some parts to operate at less than the rated frequency and b) some parts with Intel® Turbo Boost Technology 2.0 to not achieve any or maximum turbo frequencies. Performance varies depending on hardware, software, and system configuration and you can learn more at http://www.intel.com/go/turbo.

Available on select Intel® processors. Requires an Intel® HT Technology-enabled system. Your performance varies depending on the specific hardware and software you use. Learn more by visiting http://www.intel.com/info/hyperthreading.

§ Configurations: The testing was done on Based on fourth-generation Intel Xeon E5-2699 v4 @2.20 GHz processor with 22 cores, 55 MB LLC and 62 GB memory 16 1G hugepages. The testing was conducted in OPNFV Pharos testbed on Pod 12 by VSPERF community engineers

THE LINUX FOUNDATION

Acknowledgements

- Joseph Gasparakis
- Dakshina Illangovan
- Lin Yang
- Edwin Verplanke
- Priya Autee

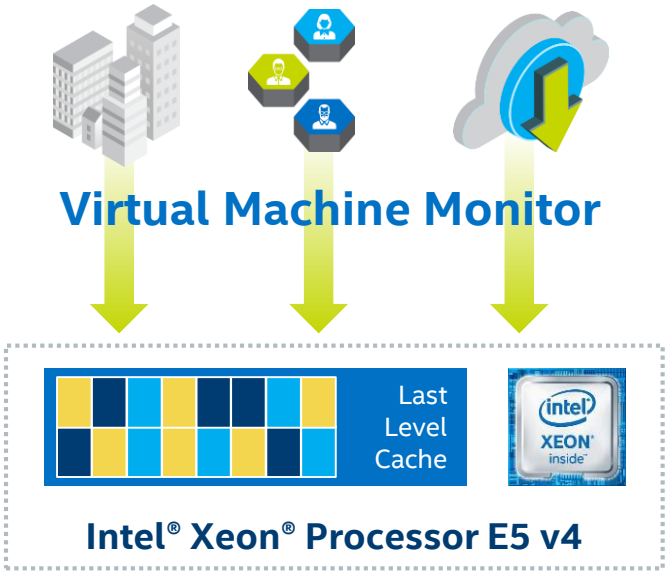
Agenda

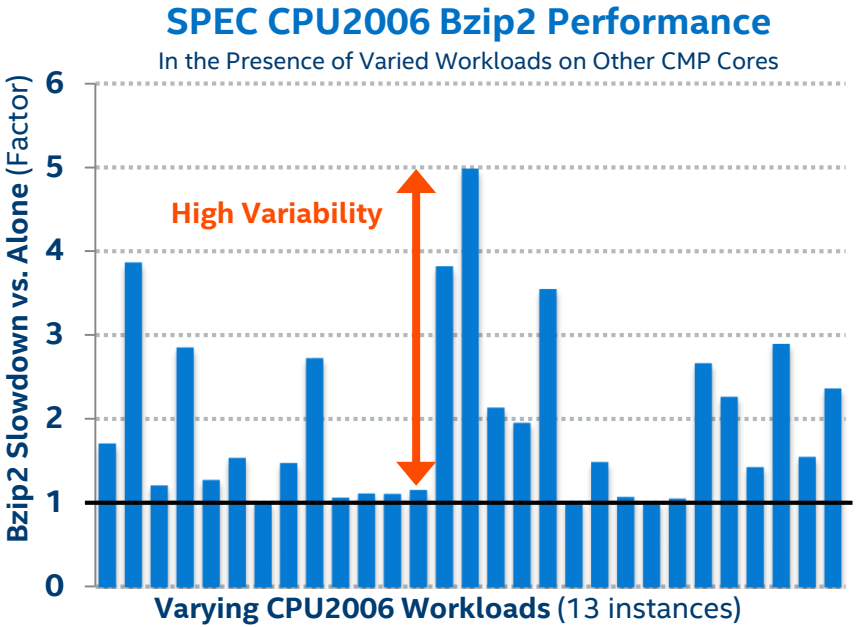
- Common Contention in Cloud
- Why is Last Level Cache Important?
- Intel Resource Director Technology
- OPNFV VSPERF, Collectd
- Resource Management Daemon
- Determinism with LLC Control

Common Contentions in Cloud Deployments

- Minimizing Total Cost of Ownership (TCO) often leads to oversubscription
- Quality of Service (QoS) requirements
  - Service Level Agreements (SLAs) Metrics: Service Availability, Throughput, Latency, Scaling.
- Cloud vs. Network Function Virtualization Deployments
  - Optimizing CPU resource utilization often leads to Shared Resource contention
- Multi-Tenants & Automated workload placement
  - Lack of control of cache by orchestration layer

Why Is Last Level Cache Important?





- Last-Level Cache Contention Can Lead to 51% Throughput Degradation[1] in Comms Workloads
- Further: Last-Level Cache Contention Can Lead to Almost 5x Performance Variation[1]

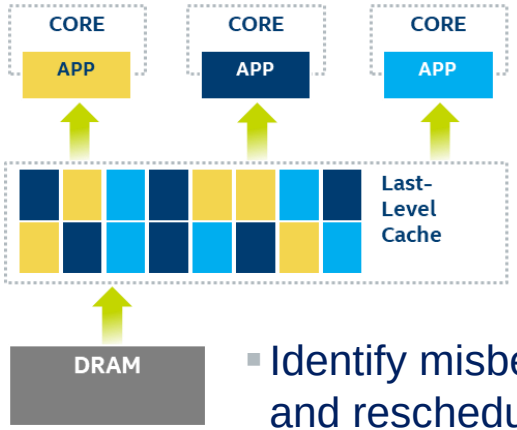
NFV & RT workloads are Time Sensitive

1: Source: UC Berkeley (UCB) Tests, 2016

THE LINUX FOUNDATION

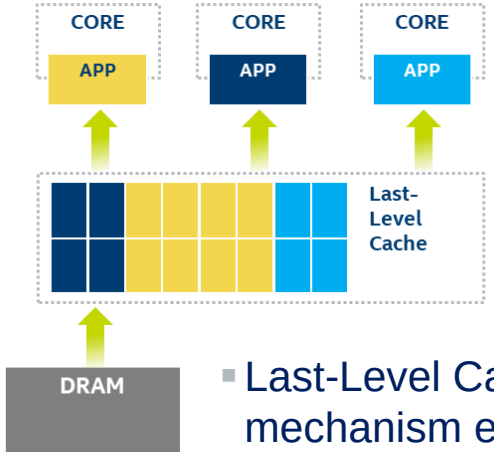
Intel® Resource Director Technology

Cache Monitoring Technology (CMT)



- Identify misbehaving applications and reschedule according to priority
- Cache Occupancy reported on a per Resource Monitoring ID (RMID) basis—Advanced Telemetry

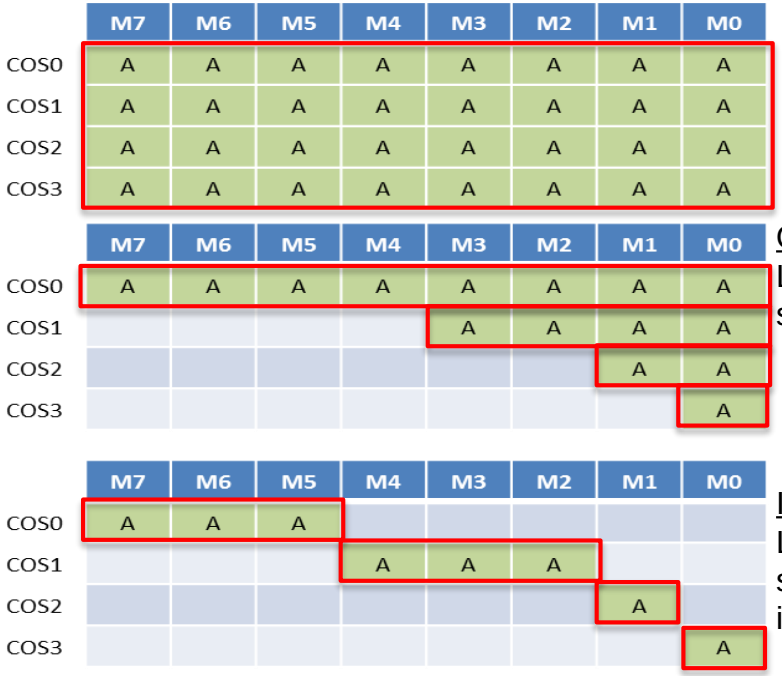
Cache Allocation Technology (CAT)



- Last-Level Cache partitioning mechanism enabling separation and prioritization of apps or VMs
- Misbehaving threads can be isolated to increase determinism

More Info: https://www.intel.com/content/www/us/en/architecture-and-technology/resource-director-technology.html

Key Concepts: Class of Service (CLOS)



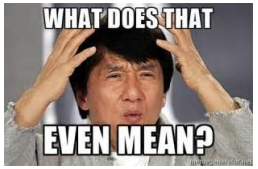
Default Bitmask
LLC is all shared

Overlapped Bitmask
LLC is partially shared.

Isolated Bitmask
LLC is allocated separately to individual COS.

- Threads/Apps/VMs grouped into Classes of Service (CLOS) for resource allocation
- Resource usage of any thread, app, VM, or a combination controlled with a CLOS
- Associate threads into CLOS
- Hardware manages resource allocation

Determinism with LLC Management



- Workload prioritization for Co-location
  - High priority, Best effort, etc.
- Cache Quality of Service (QoS) adjustments
- Consistency in Throughput & Latency
- Noisy Neighbor avoidance
  - Ex: Content Delivery Network, etc.

Impact Analysis with OPNFV VSPERF

- OPNFV VSPERF
  - Test suite to characterize the performance of a virtual switch in the NFVi
- Define, implement and execute automated test cases
- Ability to assign and scale CPUs for VNFs
- Supports multiple traffic generators and virtual switches with various VNF deployment scenarios

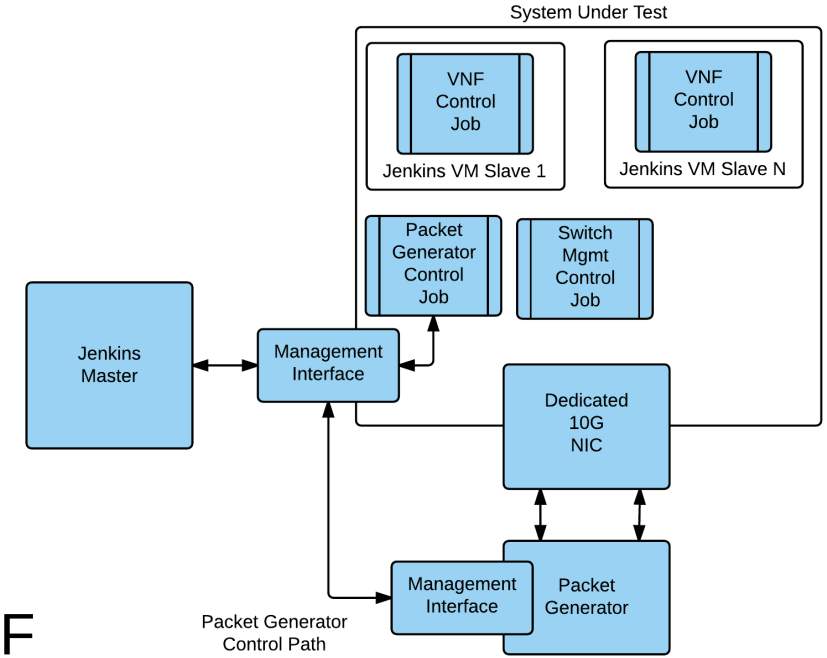


Fig: VSPERF automated test architecture

Spirent CloudStress as Noisy Neighbor

- Web-based infrastructure validation application with REST interfaces
- Emulates real-world NFV workload
- Helps performance & capacity planning for Compute, Memory, Storage & Network I/O
- Configured for heavy memory read/writes

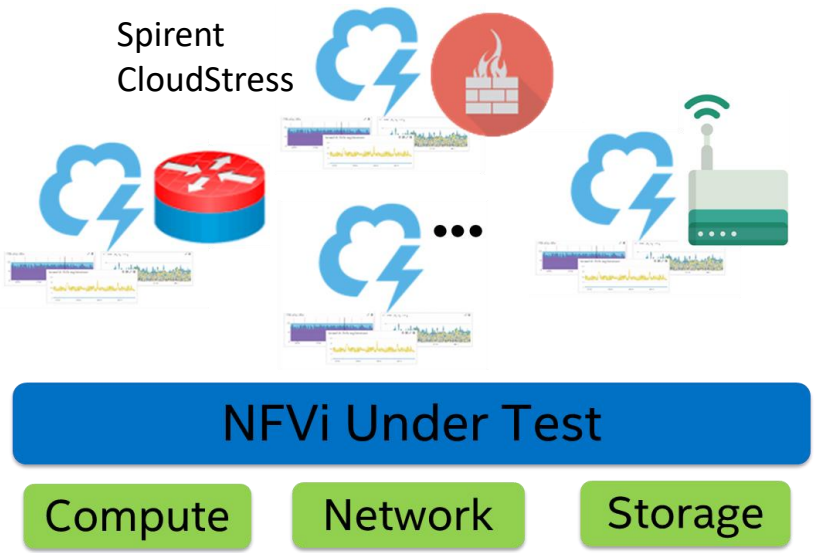


Fig: CloudStress emulates real-world VNFs

Collectd as Metrics Collector

- Statistics collection daemon
- Uses read or write plugins to collect metrics write to an end point
- Widely adopted
- Configurable collection interval
- Configurations available through OPNFV Barometer
- Leverage Intel_RDT Collectd plugin

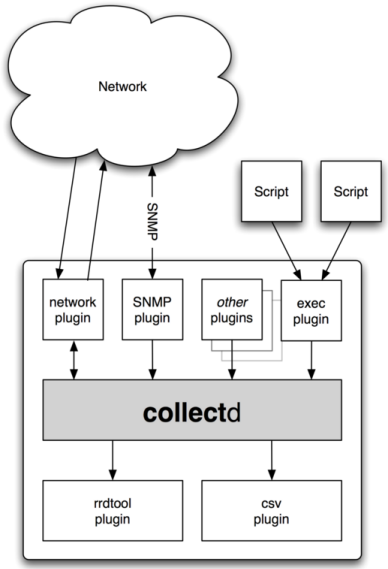


Fig: Collectd Architecture

OPNFV VSPERF Test Setup

- VSPERF integration with Collectd provides insight into NFVi data plane resource utilization
- VSPERF automates the deployment & benchmarking of NFVI setup
- L2 Forwarding VM used as VM under test

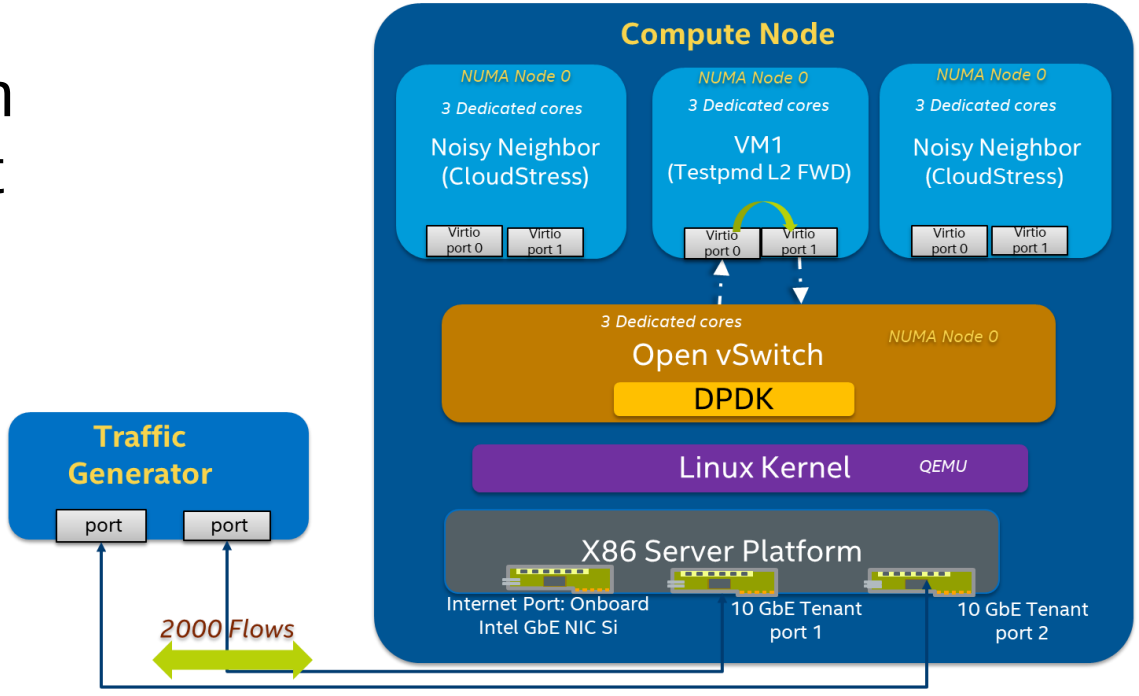


Fig: VSPERF Test Setup

THE LINUX FOUNDATION

Performance Impact with LLC Contention

- Over 33% throughput impact with Noisy Neighbor
- Heavy performance impact to the VM under test due to LLC contention

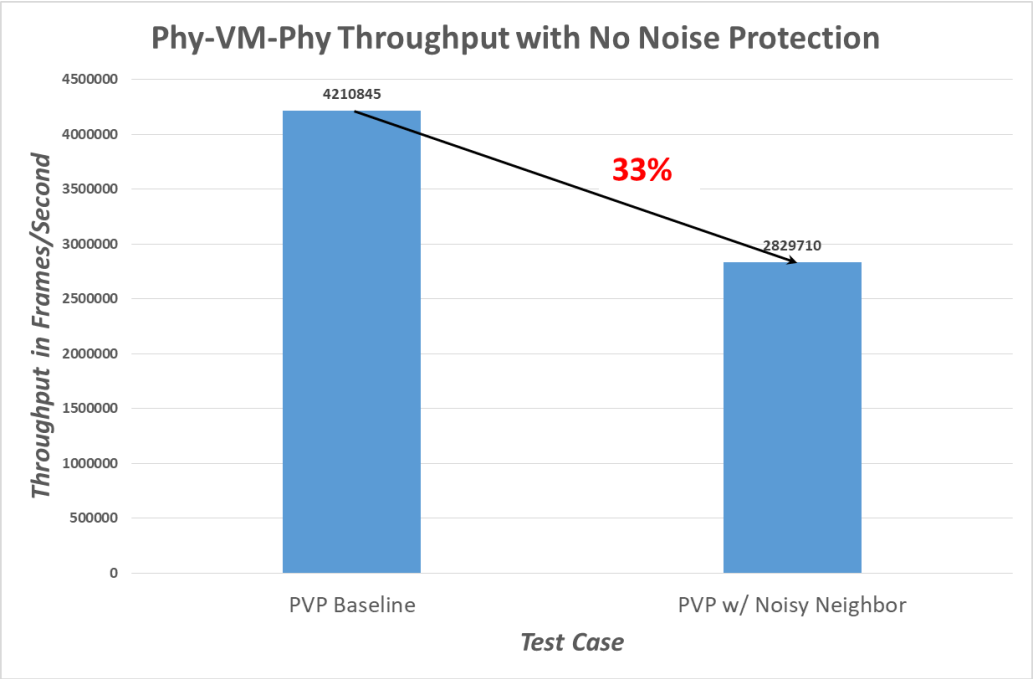


Fig: Throughput of 64B Packets

System configuration: Based on fourth-generation Intel Xeon E5-2699 v4 @2.20 GHz processor with 22 cores, 55 MB LLC and 62 GB memory 16 1G hugepages.

Class of Service Construction

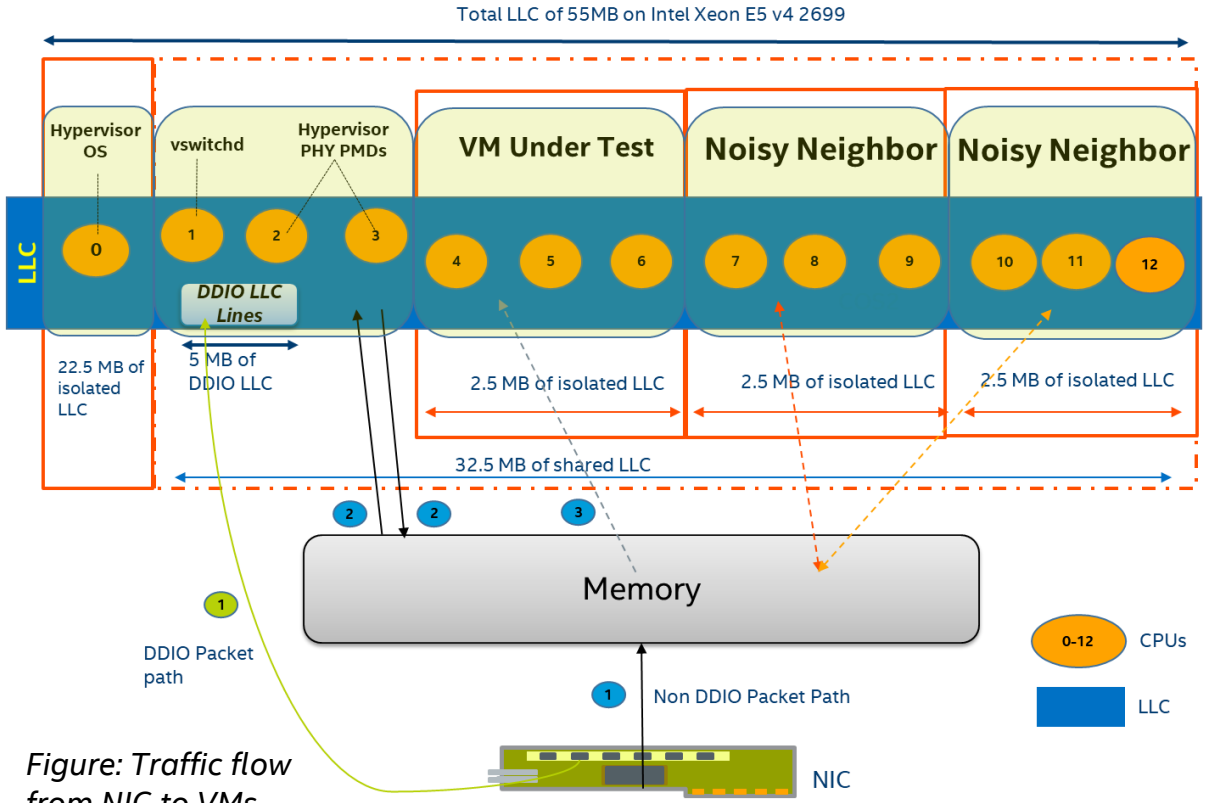


Figure: Traffic flow from NIC to VMs

Cache Profile on Intel Xeon E5-v4

- CloudStress: ~52.5MB
- vSwitchd: <2.5 MB
- DPDK PMDs: ~12.5MB/PMD
- Forwarding VM: ~2.5MB

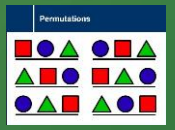
DDIO Cache: 5 MB

Optimal COS Association: OVS-DPDK overlapping VM's LLC while each VM has dedicated LLC

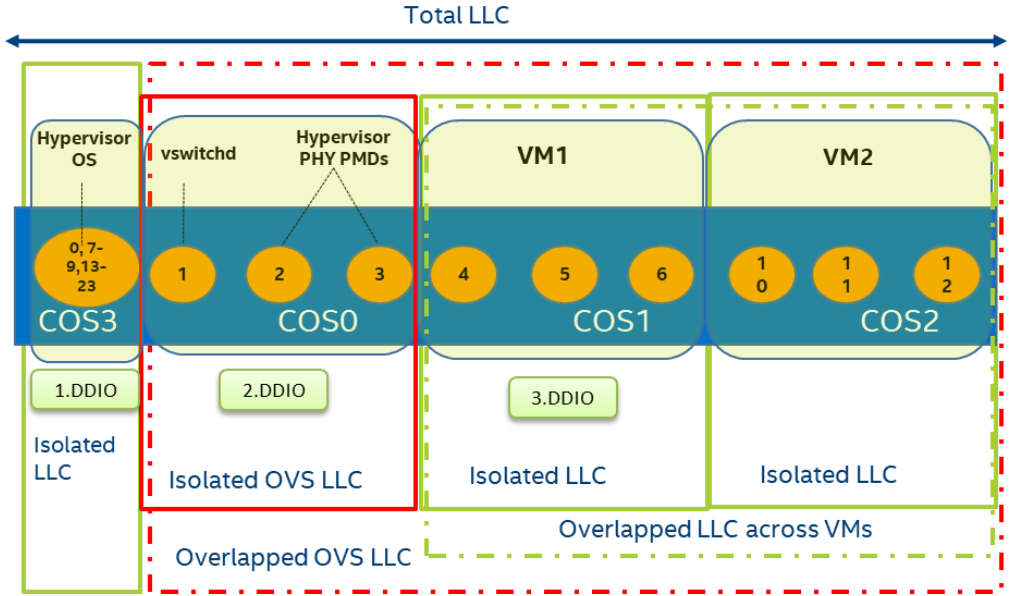
THE LINUX FOUNDATION

WhitePaper: https://builders.intel.com/docs/networkbuilders/deterministic_network_functions_virtualization_with_Intel_Resource_Director_Technology.pdf

Permutations of LLC Allocations

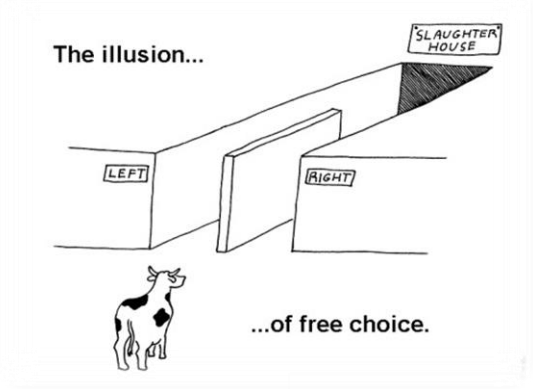


- Scheduling Considerations
  - Node capacity of cache
  - Runtime workload sensitivity and mix
- Overlapping COS between:
  - Virtual switch and VMs
  - Multiple VMs
  - OS and virtual switch
- DDIO considerations:
  - Exclusive to VMs/OS or
  - Shared across virtual switch & VMs



Planning For Resources

- Remote analysis of resource utilization and granular resource control not optimal for latency sensitive workloads
- Real time automation requires local control of LLC resources
- Planning for your Cache:
  - Translate workload requirements to policy
  - Integration with MANO Layers
  - Automated Class Of Service construction



Require Node Level Resource Manager

Resource Management Daemon (RMD)

RMD - A Linux daemon that:

- Runs on individual hosts
- REST API, accessible to orchestrator
- Accepts & enforces policy
- Platform Aware

Open Source:
https://github.com/intel/rmd

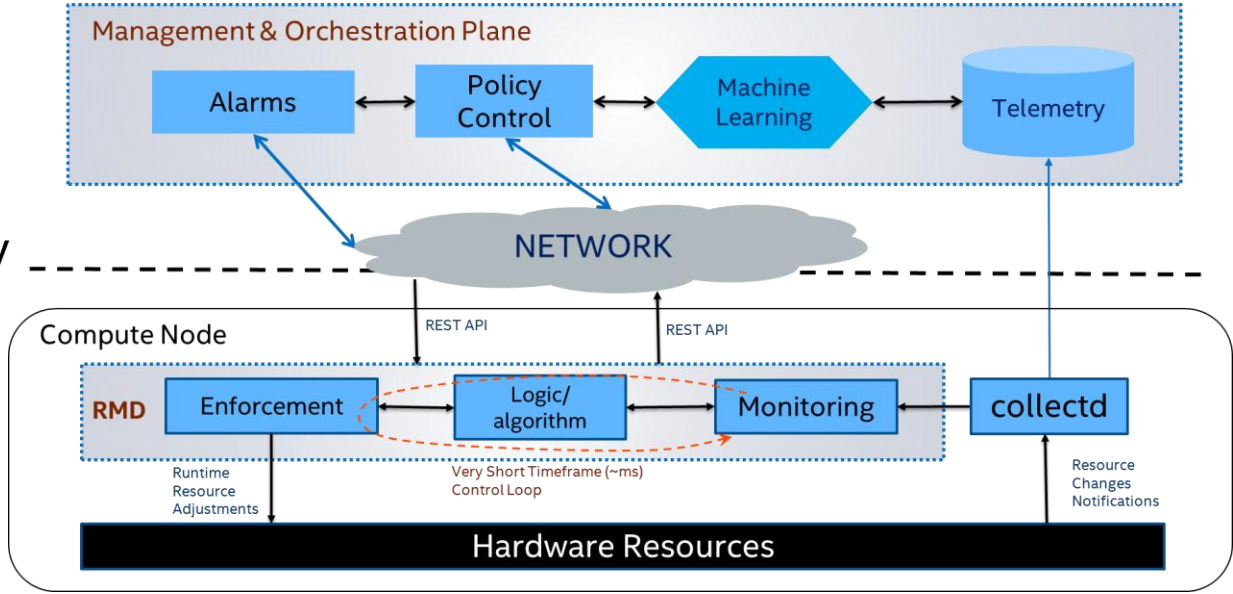


Fig: RMD interactions with Platform & MANO Layer

Why Use RMD:

- Ability to use LLC as a resource
- Satisfies multiple usecases with varying resource policies

Policy Driven LLC Allocation with RMD

- Hide COS complexity
- Pre-constructed or run time policy changes
- Scale resources at run time

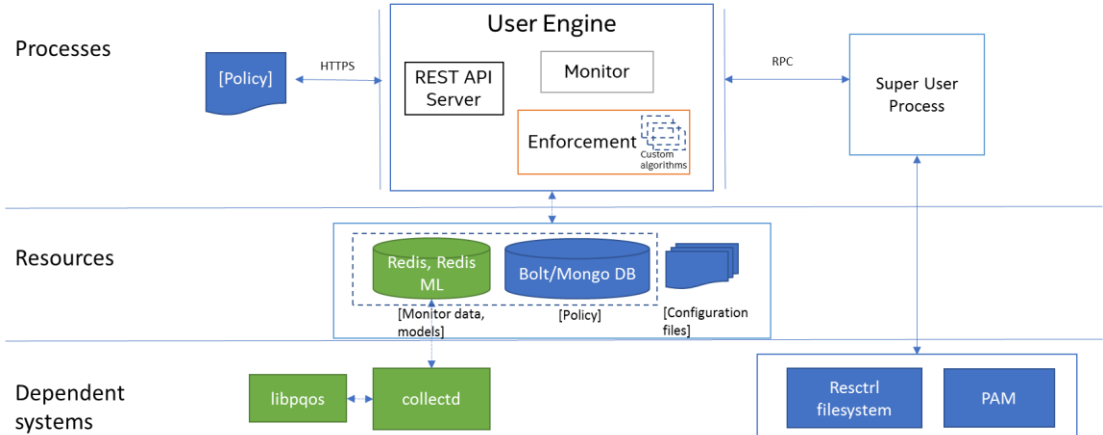


Fig: RMD Architecture

```
[CachePool]
# total 55
guaranteed = 15
burstable= 7.5
besteffort = 7.5
[OSGroup]
cache = 25
[InfraGroup]
cache = 30
```

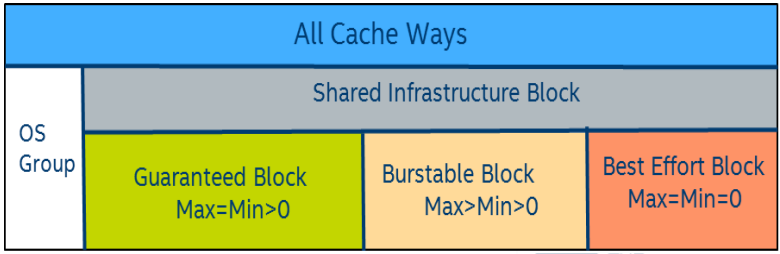


Fig: LLC Policy Customizations

THE LINUX FOUNDATION

Workload Sensitivity & Policy Mapping

- Apply LLC policy at run time using RMD
  - LLC for VM under test – “Guaranteed” - 2.5 MB
  - LLC for CloudStress VMs – “Best-effort” - 2.5 MB/VM
- Re-run the performance tests

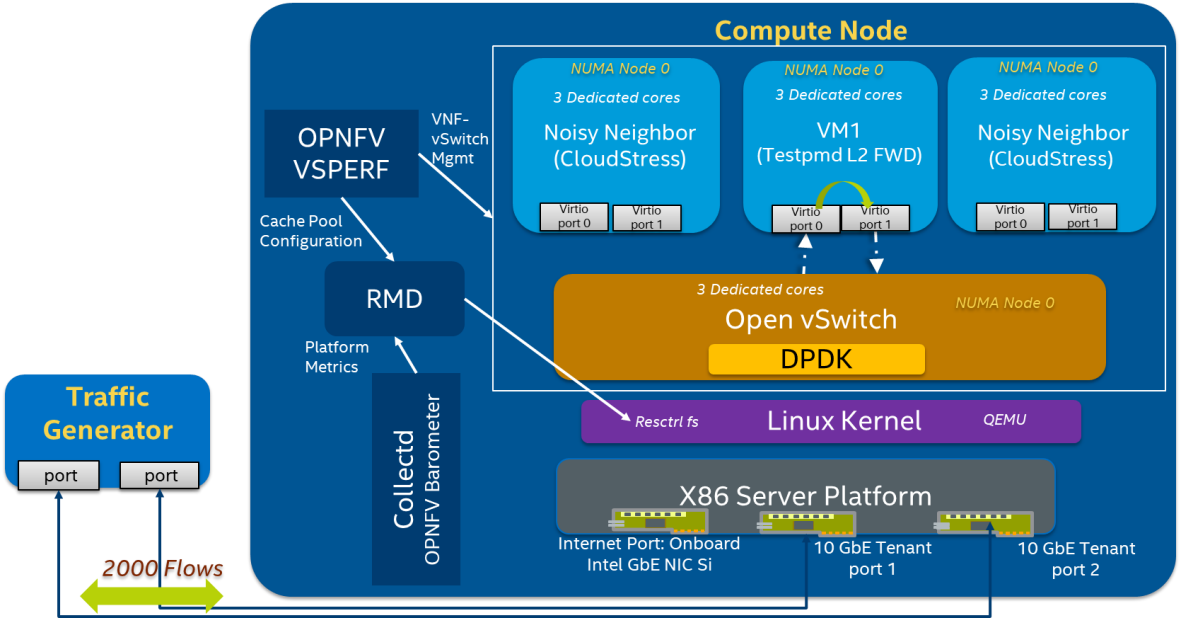


Fig: VSPERF Test setup with RMD & Collectd

Run Time LLC Control via RMD Policy Mapping

THE LINUX FOUNDATION

Noisy Neighbor Protection

- Guaranteed LLC policy helped preserve VM performance
- Throughput improvement of ~40% without noise protection

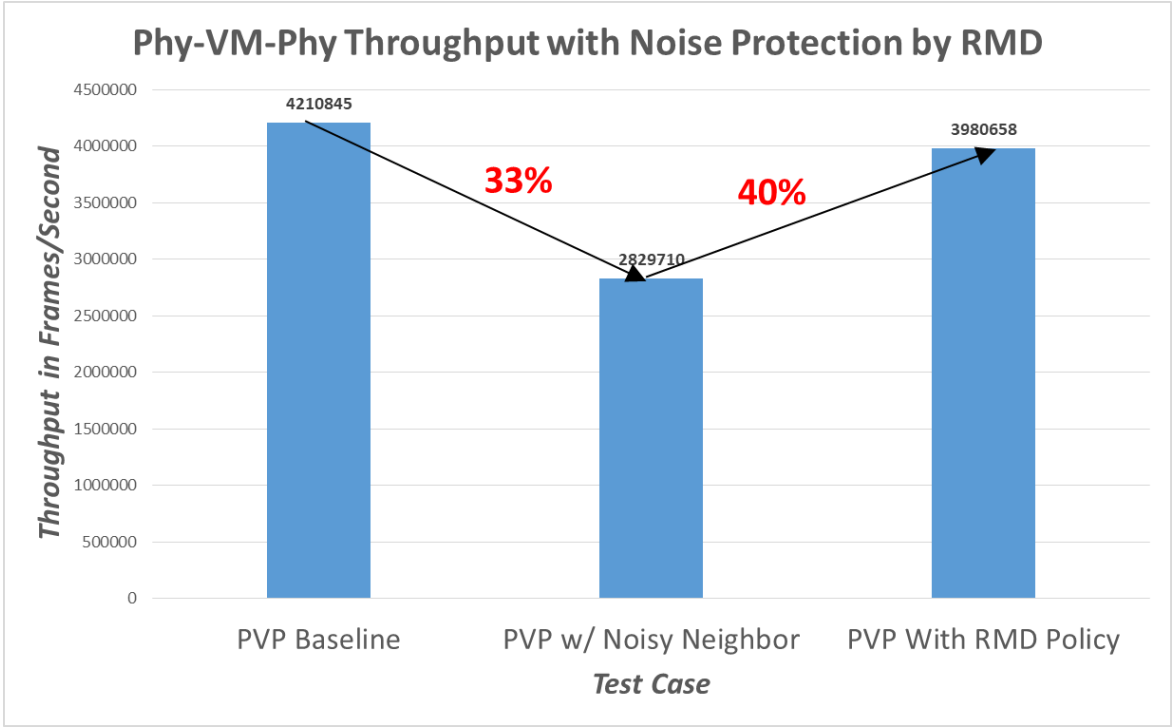


Fig: Throughput of 64B Packets

Optimal Cache Policy Ensures Deterministic Performance

THE LINUX FOUNDATION

System configuration: Based on fourth-generation Intel Xeon E5-2699 v4 @2.20 GHz processor with 22 cores, 55 MB LLC and 62 GB memory 16 1G hugepages.

Welcome to Review

- Support latency sensitive platform resources
- Integration of RMD in to OpenStack & Kubernetes
- Review blueprints/upstream work:
  - https://github.com/kubernetes/community/pull/1733
  - https://review.openstack.org/#/c/568678/6

In Summary….

- Noisy Neighbor affects are real and here to persist
- Intel Resource Director Technology enables hardware infrastructure for LLC QoS control
- RMD provides real time control of latency critical hardware resources
- OPNFV VSPERF with RMD enables LLC QoS analysis for NFVi

Update Your NFVI for LLC QoS & Control

Thank You

THE LINUX FOUNDATION

OPEN SOURCE SUMMIT

EUROPE

THE LINUX FOUNDATION