| Control Node | Control Node |

XMPP

Compute Node

vhost0

vRouter Agent
(user space)

Guest VM

vNIC

pkt0

tap

vRouter Dataplane
(Kernel Mode or DPDK)

pNIC

Title: vrouter overview	
version: 1.0	date: 25/03/2020

Compute Node

User Space

Guest VM (QEMU)

User Space

Application
(VNF)

Kernel Space

virtio-net

eth0 🖳 Virtual NIC

Nova Agent

vRouter Agent

manage

QEMU BACKEND

VirtIO

vhost

Kernel Space

bond0

pkt0

KVM

tap XXX

vif1

vif2

vifN

⊕ vrouter.ko

vRouter Dataplane

vif0

bond0

driver

🖳 🖳

Physical NICs

Title: Kernel vRouter details	
version: 1.0	date: 25/03/2020

Compute Node

User Space

Guest VM (QEMU)

User Space

Application (VNF)

virtio-net-PMD

Kernel Space

Virtual NIC

Nova Agent

vRouter Agent

manage

QEMU BACKEND

VirtIO

vhost-user

bond0

pkt0

vhost-user

vif1

vif2

vifN

vif0

DPDK Library

vRouter Dataplane

Kernel Space

KVM

PMD

Physical NICs

Title: DPDK vRouter details	
version: 1.0	date: 25/03/2020

Kernel Mode
vRouter

DPDK
vRouter

Virtio-net (Kernel)

DPDK (user)

Virtio-net (Kernel)

DPDK (user)

Compute Node

Compute Node

Compute Node

Compute Node

Guest VM

User Space

virtio-net

Kernel Space

Guest VM

User Space

virtio-net PMD

Kernel Space

Guest VM

User Space

virtio-net

Kernel Space

Guest VM

User Space

virtio-net PMD

Kernel Space

VirtIO

vhost

QEMU

VirtIO

vhost

QEMU

VirtIO

vhost-user

QEMU

VirtIO

vhost-user

QEMU

User
Space

User
Space

User
Space

User
Space

Kernel   Space

KVM

TAP

vrouter

vhost0

driver

pNIC

Kernel   Space

KVM

TAP

vrouter

vhost0

driver

pNIC

vhost-**user**

**DPDK
vrouter**

vhost0

Kernel
Space

PMD

pNIC

vhost-**user**

**DPDK
vrouter**

vhost0

Kernel
Space

PMD

pNIC

Title: Kernel vRouter All Cases	
version: 1.0	date: 25/03/2020

Core

Core

VM VirtIO
Interface

NUMA 0

QPI

NUMA 1

Physical
Interface

Physical
Interface

Core

Core

VM VirtIO
Interface

NUMA 0

QPI

NUMA 1

VM VirtIO
Interface

Physical
Interfaces

Title: NUMA concern	
version: 1.0	date: 25/03/2020

CPU Pinning

VROUTER	
VNF	
Operating System Hypervisor (QEMU/KVM)	
VROUTER DPDK initial setup	

Process to de scheduled

- Forwarding Threads (lcores >= 10)
- Service Threads (lcores <= 9)
- VNF instance X vCPU
- VNF instance Y vCPU
- VNF instance Z vCPU
- Operating System Processes
- Control Threads

CPU mapping Configuration

- vrouter CPU_LIST (0x154)
- vrouter SERVICE_CORE_MASK (0x400 000000 400)
- NOVA - vcpu_pin_set 3,5,7,9,11-35,39,41,43,45,47-71
- Systemd CPUAffinity 0,1,36,37
- vrouter DPDK_CTRL_TREAD_MASK (0x003 000000 003)

Scheduling Mechanism

- Taskset = 2 or 4 or 6 or 8
- vrouter process option service_core_mask (0x400 000000 400)
- LIBVIRT cpuset « X »
- LIBVIRT cpuset « Y »
- LIBVIRT cpuset « Z »
- Systemd
- vrouter process option dpdk_ctrl_tread_mask (0x003 000000 003)

tuned.isolated_cores + GRUB isolcpus

Systemd.CPUAffinity

Physical Cores

2 4 6 8 | 10 | 12 ... 34 | 3 5 7 ... 35 | 0 | 1

Assigned for vrouter packet processing

Allocated for NOVA (VNF)

Assigned for generic purpose

Sibling Cores

38 40 42 44 | 46 | 48 ... 70 | 39 41 43 ... 72 | 36 | 37

NUMA 0 | *NUMA 1* | *NUMA 0* | *NUMA 1*

Application CPUS | **System CPUS**

Title: CPU Pinning	
version: 1.0	date: 25/03/2020

mbuf

MEMPOOL (set of MBUFs)

Packet
Transfer

memory
pointer

QUEUE pkt

Desc.
Ring

**LCORE
(pthread)**

Title: DPDK High Level Components	
version: 1.0	date: 25/03/2020

vrouter Queues size:
--vr_dpdk_tx_ring_sz (1024)
--vr_dpdk_rx_ring_sz (1024)

Allow or not forwarding cores to be released:
--yield_option (1)

Allow or not core re-balancing (enable or not internal hasing):
TBD

Physical NIC Queues size:
--dpdk_txd_sz (128)
--dpdk_rxd_sz (128)

vRouter

Polling Cores

Processing Cores

vrouter hash

DPDK RSS hash

Physical NIC

Device Ring

Device Ring

Device Ring

Device Ring

LCORE pthread

LCORE pthread

LCORE pthread

LCORE pthread

Intern. Q

Intern. Q

Intern. Q

Intern. Q

LCORE pthread

LCORE pthread

LCORE pthread

LCORE pthread

VirtIO Ring

VirtIO Ring

VirtIO Ring

VirtIO Ring

VM DPDK Application

Virtual NIC

« vif 0/0 »
Physical NIC Queues

« vif 0/N »
vNIC Queues

Service & Control Cores

LCORE pthread

LCORE pthread

Mempool size
--vr_mempool_sz (16384)

Service and Control cores pinning:
--service_core_mask
--dpdk_ctrl_thread_mask

MEMPOOL (DPDK rings memory)

Title: vRouter Architecture	
version: 1.0	date: 25/03/2020

Polling Core

Processing Core

DPDK RSS hash

Device Ring

LCORE (pthread)

Intern. Q

LCORE (pthread)

Processing Cores Are writing one by one Into the single Q

Device Ring

LCORE (pthread)

Intern. Q

LCORE (pthread)

No Hash recalculation Polling Core = Processing Core

Device Ring

LCORE (pthread)

Intern. Q

LCORE (pthread)

Device Ring

LCORE (pthread)

Intern. Q

LCORE (pthread)

VirtIO Ring

Physical NIC

« vif 0/0 » Physical NIC Queues

vRouter processing Queues

« vif 0/N » vNIC Queue

Virtual NIC

VM DPDK Application

Title: UDP to SingleQ	
version: 1.0	date: 25/03/2020

Polling Core

vrouter hash

Device Ring

LCORE (pthread)

Intern. Q

LCORE (pthread)

Device Ring

LCORE (pthread)

Intern. Q

LCORE (pthread)

Device Ring

LCORE (pthread)

Intern. Q

LCORE (pthread)

Device Ring

LCORE (pthread)

Intern. Q

LCORE (pthread)

VirtIO Ring

Processing Cores Are writing one by one Into the single Q

DPDK RSS hash

Physical NIC

« vif 0/0 » Physical NIC Queues

vRouter processing Queues

« vif 0/N » vNIC Queue

Virtual NIC

VM DPDK Application

Title: GRE to SingleQ	
version: 1.0	date: 25/03/2020

Polling Core

Processing Core

vrouter hash

DPDK RSS hash

Only one core is polling the single Q

LCORE pthread

LCORE pthread

LCORE pthread

LCORE pthread

Intern. Q

Intern. Q

Intern. Q

Intern. Q

LCORE pthread

LCORE pthread

LCORE pthread

Device Ring

Device Ring

Device Ring

Device Ring

VirtIO Ring

Virtual NIC

Physical NIC

« vif 0/N »
vNIC Queues

vRouter
processing Queues

« vif 0/0 »
Physical NIC Queues

Title: SingleQ to Underlay	
version: 1.0	date: 25/03/2020

Polling Core

Processing Core

DPDK RSS hash

Device Ring

LCORE (pthread)

Intern. Q

LCORE (pthread)

VirtIO Ring

No Hash recalculation
Polling Core = Processing Core

Physical NIC

« vif 0/0 »
Physical NIC Queues

vRouter
processing Queues

« vif 0/N »
vNIC Queues

VM
DPDK
Application

Virtual NIC

Title: UDP to MultiQ	
version: 1.0	date: 25/03/2020

Polling Core

Processing Core

vrouter hash

DPDK RSS hash

Device Ring

LCORE (pthread)

Intern. Q

LCORE (pthread)

VirtIO Ring

Physical NIC

« vif 0/0 »
Physical NIC Queues

vRouter processing Queues

« vif 0/N »
vNIC Queues

VM DPDK Application

Virtual NIC

Title: GRE to MultiQ	
version: 1.0	date: 25/03/2020

Polling Core

Processing Core

vrouter hash

DPDK RSS hash

VirtIO Ring

VirtIO Ring

VirtIO Ring

VirtIO Ring

LCORE pthread

LCORE pthread

LCORE pthread

LCORE pthread

Intern. Q

Intern. Q

Intern. Q

Intern. Q

LCORE pthread

LCORE pthread

LCORE pthread

LCORE pthread

Device Ring

Device Ring

Device Ring

Device Ring

Virtual NIC

Physical NIC

« vif 0/N » vNIC Queues

vRouter processing Queues

« vif 0/0 » Physical NIC Queues

Title: MultiQ to Underlay	
version: 1.0	date: 25/03/2020

vif 0
RX device packets
RX port packets

vif 0
RX Queue packets
RX Queue errors
TX Queue packets
TX Queue errors

vif 0
RX packets
vif N
TX packets

vif N
TX port packets

vRouter

Polling Cores

Processing Cores

Device Ring

LCORE (pthread)

Intern. Q

LCORE (pthread)

VirtIO Ring

DPDK RSS hash

Physical NIC

« vif 0/0 »
Physical NIC Queues

« vif 0/N »
vNIC Queues

VM DPDK Application

Virtual NIC

Title: vif counters (phys to virt)	
version: 1.0	date: 25/03/2020

vif N
RX port packets

vif N
RX Queue packets
RX Queue errors
TX Queue packets
TX Queue errors

vif N
RX packets

vif 0
TX packets

vif 0
TX port packets

vRouter

Polling Core

Processing Core

vrouter hash

DPDK RSS hash

VirtIO Ring

LCORE pthread

Intern. Q

LCORE pthread

Device Ring

Virtual NIC

« vif 0/N »
vNIC Queues

vRouter processing Queues

« vif 0/0 »
Physical NIC Queues

Physical NIC

Title: vif counters (virt to phys)	
version: 1.0	date: 25/03/2020

7 CPU DPDK VNF

Core Core Core Core Core Core Core

V I R T I O

7 Q R I N G S

Core Core Core Core

4 CPU VROUTER

COMPUTE NODE

VM VM VM VM VM VM

vRouter (DPDK)
4 cores

bond   vlan X

SRIOV NIC   SRIOV NIC

TOR Switch

---

mempool for mbufs (64k)

mbuf

pkt0
(vRouter agent)
vif0/2

CPU core 0 vrouter

CPU core 1 vrouter

CPU core 2 vrouter

CPU core 3 vrouter

QEMU
DPDK VM
1Q virtual core 1
vif0/3

QEMU
non-DPDK VM
1Q virtual core 1
vif0/4

Intel NIC 82599

RX QUEUE (core 0)
pkt
vif0/0

vHost0
(host kernel)
vif0/1

1   2

---

mempool for mbufs (64k)

mbuf

pkt0
(vRouter agent)
vif0/2

CPU core 0 vrouter

CPU core 1 vrouter
processing core

CPU core 2 vrouter

CPU core 3 vrouter
polling core

QEMU
DPDK VM
1Q virtual core 1
vif0/3

QEMU
non-DPDK VM
1Q virtual core 1
vif0/4

Intel NIC 82599

TX QUEUE (core 1)
vif0/0

vHost0
(host kernel)
vif0/1

1   2

---

mempool for mbufs (64k)

mbuf

pkt0
(vRouter agent)
vif0/2

CPU core 0 vrouter
polling core

CPU core 1 vrouter
processing core

CPU core 2 vrouter

CPU core 3 vrouter

QEMU
DPDK VM
1Q virtual core 1
vif0/3

QEMU
non-DPDK VM
1Q virtual core 1
vif0/4

Intel NIC 82599

RX QUEUE (core 0)
pkt
vif0/0

vHost0
(host kernel)
vif0/1

1   2

---

vif0/0

TX queue 0   TX queue 1   TX queue 2   TX queue 3

Desc. Ring   Desc. Ring   Desc. Ring   Desc. Ring

mempool for mbufs (64k)

LCORE (pthread)   LCORE (pthread)   LCORE (pthread)   LCORE (pthread)

vif0/N

pNIC TX queue

Physical NIC Memory

Device Ring

vNIC TX queue 0   VirtIO Ring

vNIC TX queue 1   VirtIO Ring

vNIC TX queue 2   VirtIO Ring

vNIC TX queue 3   VirtIO Ring

VM Memory

Virtual NIC

Virtual NIC (virtio backend)   Virtual NIC (virtio frontend)   DPDK Application

Physical NIC   **dpdk vRouter**   **QEMU/KVM**   **Virtual Machine**

Title: Personal Stuff	
version: 1.0	date: 25/03/2020

VM 0 · VM 1 · DPDK PMD
virtio FE · Qemu · rte_ring · KVM · Linux · User · Kernel · vring · vHost-net · KNI Module
Intel Architecture · DEV (Niantic)

VirtIO · VirtIO · vhost-user · vhost user · Kernel · NIC

QEMU Operating System · Virtio Driver · DPDK vhost · DPDK x · IOCTL · CUSE · KVM · eventfd link · ioeventfd · irqfd · Kernel Space · User Kernel

Guest 1 · Guest 2 · Application · Userspace · Virtio-net Driver · Kernel · vhost_user · DPDK · (bypassed) · Kernel · ixgbe · qemu-kvm · Guest · Host · User space

KERNEL MODE · DPDK MODE · SRIOV (vrouter bypassed)
VM · Hypervisor · Server NIC · PHY · DPDK Lib · VM

PMD CPU#0 · VROUTER polling lcore · queue#0 · vhost-user q#0 · VM · vCPU #0 · virtio-net queue #0
PMD CPU#1 · queue#1 · vhost-user q#1 · vCPU #1 · queue #1
PMD CPU#2 · queue#2 · vhost-user q#2 · vCPU #2 · queue #2
PMD CPU#4 · queue#3 · vhost-user q#3 · vCPU #3 · queue #3
10GbE · NIC RSS

phys. memory: hugepages · free
rte_memzone_reserve(name, len, socket, size)
memory zones: ring zone · mempool zone · free · heap zone
rte_mempool_create(name, n, elt_size, ...)
rte_mempool: free mbufs · mbuf · free · mbuf · free · mbuf · mbuf · mbuf · free · private · next
rte_pktmbuf_alloc(mempool)
rte_pktmbuf: priv · headroom · data · tailroom

Server · NUMA SOCKET 0 · NUMA SOCKET 1 · CPU 0 · QPI-0 · QPI-1 · RAM 0 · RAM 1 · IO Controller · PCIe slots · PCIe0 · PCIe1 · NICs · TOR switch

User space · App · DPDK · Ring buffers
Kernel space · UIO driver
NIC · RX/TX queues

lcore 1 · TX ring · RX ring · memory zones: ring zone · mempool zone · free · heap zone · rte_mempool_create(name, n, elt_size, ...) · rte_mempool: free mbufs · mbuf · free · mbuf · free · mbuf · mbuf · mbuf · free · private

Incoming traffic · Hash function · Indirection table · Queue 0 · ... · Queue N · CPU N

Memory · Descriptor ring · RX queue · RX FIFO · Update RDT · DMA descriptor(s) · DMA packet · DMA descriptors · Host memory · NIC memory
Memory · Descriptor ring · TX queue · TX FIFO · Update TDT · DMA descriptor(s) · DMA packet · DMA descriptors · Host memory · NIC memory

User Space · QEMU Layer · VM · Nova Agent · Guest VM · Application VM DPDK · eth0 · eth1 · User Space · VIRTIO · Kernel Space · vRouter Host Agent · vRouter (VRFWD) · Kernel Space

Computer Node · VM1 · VM2 · vrouter-dpdk · c1 · c2 · c3 · c4 · NIC

Main Memory · Physical addresses · IOMMU · MMU · Device addresses · Virtual addresses · Device · CPU

Host · Guest: 4 vCPUs · Guest: 1 vCPU · Virtio DPDK PMD · VirtIO DPDK PMD · CPU0 · CPU1 · CPU2 · CPU3 · user land · kernel · vhost-user · vRouter DPDK · user land · kernel

VHOST client · shared memory · socket · VHOST Backend

Host Compute Node · vhost-net : Kernel Space (Before QEMU 2.1) · vhost-User: User Space vhost (QEMU 2.1) · Host Process per VM · Guest VM (Virtual Machine) · VM · QEMU 2.2 Process Per VM · Qemu Uvhost client · vRouter Agent · VIRTIO Ring · DPDK 2.0 Libraries · User-Space · Kernel Space · User Space vhost (libvirt 1.2.7) · Config Policy Tables · Poll · DPDK lcores · Virtio · VRFWD hugetlbfs (DPDK Ring) · VIRTIO Backend · netlink · Uvhost Server · vRouter Forwarding · pkt0 · DPDK NIC · NIC Queues (1,2..N) · Lcores to NIC Queue Mapping 1-1 · DPDK vRouter

NUMA 0 · Host OS · DPDK PMDs · Memory · VMs · cpu-mask=0xF0 · socket-mem=1024,1024
NUMA 1 · Host OS · DPDK PMDs · Memory · VMs

(1) Single queue (Previous result) · (2) Multi queue (Next result) · Guest · vNIC · Fast-Path · cX vcpu · qX Queue

Title: Stuff (no source) · version: 1.0 · date: 25/03/2020