

***TF vRouter/DPDK Performance Report***

***Release R2003***

**Authors: Kiran KN <**[**kirankn@juniper.net**](mailto:kirankn@juniper.net)**>,**

**Przemysław Grygiel <**[**pgrygiel@juniper.net**](mailto:pgrygiel@juniper.net)**>,**

**Damian Szeluga <**[**dszeluga@juniper.net**](mailto:dszeluga@juniper.net)**>**

***Revision History***

|  |  |  |
| --- | --- | --- |
| Date | Revision | Comment |
| April 6th 2020 | 1.0 | Initial document |
| May 2020 | 2.0 | Final document |

***Contents***

[***Audience and Purpose***](#_m9d1ebg76i9p) ***7***

[***Hardware Specifications***](#_krsqemp5zw4r) ***7***

[*General CPU allocation*](#_6vuksam9ww9d) *8*

[*Hypervisor configuration description*](#_nnmv04tgjrwp) *8*

[***Test setup***](#_f88w6m5zfvkg) ***9***

[***About PROX and Rapid***](#_cif8uuxb45j5) ***10***

[*Prox VMs tuning description (compute-B)*](#_8zxmqjwfza9h) *10*

[***Test case 1: Number of packets per second (Mpps) and Throughput (Gbps) per core on Intel 82599 Niantic Ethernet Adapters***](#_ftv97zvm9ip6) ***12***

[*Configuration of vRouter on compute-B (DUT):*](#_2fgnu73x2c4c) *12*

[*CPU allocation (Full-CPU partitioning scheme) on compute-B:*](#_odwgoab97mtn) *12*

[*Traffic pattern*](#_ni0607md99f4) *12*

[*Packet mode (Mpps):*](#_21kgjrgs6k1a) *13*

[*Packet mode (Gbps):*](#_nfjvk7gwbqf0) *14*

[*Flow mode (Mpps):*](#_qzn97qpa1ang) *15*

[*Flow mode (Gbps):*](#_rm03k85h0dct) *16*

[*Summary*](#_wo0yt3f228cx) *17*

[*82599 Niantic configuration*](#_9j5nlffhwqib) *17*

[*Detailed data*](#_77vwlqww0lh5) *18*

[***Test case 2: Number of packets per second (Mpps) and Throughput (Gbps) per core on Intel X710 Fortville Ethernet Adapters***](#_8j5mafsi3whj) ***19***

[*Configuration of vRouter on compute-B (DUT):*](#_q00etu3qtjfo) *19*

[*CPU allocation (Full-CPU partitioning scheme) on compute-B:*](#_gei9qb741cvg) *19*

[*Traffic pattern*](#_p9jihiu9kw59) *19*

[*Packet mode (Mpps):*](#_25rm2nx1eg5l) *20*

[*Packet mode (Gbps):*](#_zbxqyrpyqeol) *21*

[*Flow mode (Mpps):*](#_pd2ak1uki87m) *22*

[*Flow mode (Gbps):*](#_9ani92rd2eyj) *23*

[*Summary*](#_bxky2nwbpgbi) *24*

[*Intel x710 Fortville configuration*](#_s8dqbg2opz9p) *24*

[*Detailed data*](#_m9k6ymswluc2) *25*

[***Test case 2a: Number of packets per second (Mpps) and Throughput (Gbps) Intel X710 Ethernet Adapters - low latency***](#_cfhdxpikz2vh) ***26***

[*Configuration of vRouter on compute-B (DUT):*](#_4tkjciyy2sy) *26*

[*CPU allocation (Full-CPU partitioning scheme) on compute-B:*](#_vujfmf19fank) *26*

[*Traffic pattern*](#_5d2k9l162pq4) *26*

[***Packet mode***](#_8a5pwfps6mz3) ***27***

[***Flow mode***](#_bfiurwxwm860) ***29***

[*Summary*](#_spvf88i79y51) *31*

[*Intel x710 Fortville configuration*](#_b4991a24teim) *31*

[*Detailed data*](#_ws1b1c2ko23v) *31*

[***Test case 3: Number of packets per second (Mpps) and Throughput (Gbps) per core on Intel XXV710 Ethernet Adapters***](#_fav9aczi64ex) ***33***

[*Configuration of vRouter on compute-B (DUT):*](#_jx9xrdoc7dbk) *33*

[*CPU allocation (Full-CPU partitioning scheme) on compute-B:*](#_y05ou01915nj) *33*

[*Traffic pattern*](#_r337q23odn3w) *33*

[*Packet mode (Mpps):*](#_1ok6u67z9dyz) *34*

[*Packet mode (Gbps):*](#_oyj9by4fhzff) *35*

[*Flow mode (Mpps):*](#_yw6bj29pzf) *36*

[*Flow mode (Gbps):*](#_2v4cjndf0ykd) *37*

[*Summary*](#_tj7lzd6btvu4) *38*

[*Intel XXV710 Fortville configuration*](#_u45y07dy6icc) *38*

[*Detailed data*](#_3weuwm167fip) *39*

[***Test case 4: Compare NICs results Intel 82599 vs Intel X710 vs Intel XXV710***](#_zae7j0gk3fab) ***40***

[*Configuration of vRouter on compute-B (DUT):*](#_62j9p5gcg95j) *40*

[*CPU allocation (Full-CPU partitioning scheme) on compute-B:*](#_2dussr7myvfw) *40*

[*Traffic pattern*](#_tgh2fkw4jnta) *40*

[*Packet mode:*](#_g08dl5ti89nq) *41*

[*Flow mode:*](#_g9r001y1klig) *42*

[*Detailed data*](#_jdz61ir7wq83) *43*

[***Test case 5: Number of packets per second (Mpps) and Throughput (Gbps) per core with increasing number of vRouter cores***](#_9fj1w65pn8m) ***44***

[*Configuration of vRouter on compute-B (DUT):*](#_ykii27g6xs7o) *44*

[*CPU allocation (Full-CPU partitioning scheme) on compute-B:*](#_1ljcama9l2b3) *44*

[*Traffic pattern*](#_mb0zhj4vbjk0) *44*

[*Packet mode without siblings (HT):*](#_e4961p5d0llj) *45*

[*Packet mode with siblings (HT):*](#_pidnjoen245s) *46*

[*Flow mode without siblings (HT):*](#_k5a0943eiwmc) *47*

[*Flow mode with siblings (HT):*](#_mk0icz147f9q) *48*

[***Test case 6: Number of packets per second (Mpps) per core with increasing number of flows***](#_zhymoiq9a5v) ***50***

[*Configuration of vRouter on compute-B (DUT):*](#_6h5myod2pu5q) *50*

[*CPU allocation (Full-CPU partitioning scheme) on compute-B:*](#_tnhv1onl8x17) *50*

[*Traffic pattern*](#_hgxwce6dljdz) *50*

[*Packet mode:*](#_d3ilqf5yby36) *51*

[*Flow mode:*](#_xsojr9gvqb7l) *51*

[*Intel x710 Fortville configuration*](#_cl67c98srt1b) *52*

[***Test case 7: Packets per second (Mpps) with different NUMA placement for VM***](#_svy1ss5799ek) ***53***

[*Configuration of vRouter on compute-B (DUT):*](#_thpcwp6gljh4) *53*

[*CPU allocation (Full-CPU partitioning scheme) on compute-B:*](#_7dmqyu76s61x) *53*

[*Traffic pattern*](#_3rqecslqp8ad) *53*

[*Packet mode:*](#_7l0rby2hcuqt) *54*

[*Flow mode:*](#_x1s7e6qlbf7h) *54*

[*Summary*](#_d70vaxryvt97) *55*

[*Intel x710 Fortville configuration*](#_polajlwet91g) *55*

[***Test case 8: Number of packets per second (Mpps) per core with different HugePage sizes***](#_aqr9svs1plc2) ***56***

[*Configuration of vRouter on compute-B (DUT):*](#_oo6wugrmmr44) *56*

[*CPU allocation (Full-CPU partitioning scheme) on compute-B:*](#_ix6ud3tfge5v) *56*

[*Traffic pattern*](#_5wc0a3lo9roa) *56*

[*Packet mode:*](#_dk97yulkpkcl) *57*

[*Flow mode:*](#_kpatf69tqo9o) *57*

[*Summary*](#_b9rack9qsaym) *58*

[*Intel x710 Fortville configuration*](#_y8efkqe0l9ln) *58*

[***Appendix***](#_2sx2ytwnj0n0) ***59***

[*Guide how to run testing framework*](#_syami55yelqj) *59*

[*Configure high performance Contrail Cloud computes*](#_36h9ima8mi8t) *59*

[*Configure high performance Contrail Networking on computes with TripleO heat templates*](#_msd6p6okbczn) *62*

[***References***](#_cgule3kkneik) ***66***

# ***Audience and Purpose***

The primary audience for this test report are system engineers, sales engineers and solution architects looking at the dataplane of Tungsten Fabric. This report presents information about the packet processing performance for the specified release on the specified testbed.

The purpose of this report is to provide guidance in terms of base line numbers that one can expect if they follow the same set of steps on a similar testbed with similar hardware configuration. This will help guide architects and engineers who are evaluating and implementing dataplane solutions and can assist in achieving optimal system performance.

# ***Hardware Specifications***

|  |  |
| --- | --- |
| Model name | Intel(R) Xeon(R) Gold 5120 CPU @ 2.20GHz |
| CPUs | 56 |
| Sockets | 2 |
| Cores per Socket | 14 |
| Threads per core | 2 |
| NUMA 0 CPU list | 0,2,4,6,8,10,12,14,16,18,20,22,24,26,  28,30,32,34,36,38,40,42,44,46,48,50,52,54 |
| NUMA 1 CPU list | 1,3,5,7,9,11,13,15,17,19,21,23,25,27,  29,31,33,35,37,39,41,43,45,47,49,51,53,55 |

## 

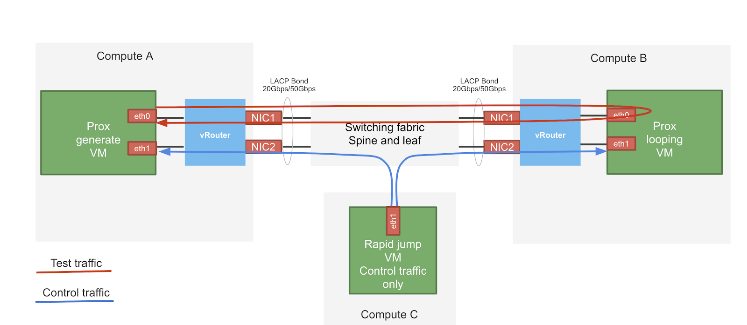
## General CPU allocation

|  |  |
| --- | --- |
|  | Core list |
| Nova cores allocation for VMs | NUMA 0: 10,12,14,16,18,20,22,24,26,38,40,42,44,46,48,50,52,54  In some test also cores from NUMA1: 1,3,5,7,9,11,13,15,17,19,21,23,25,27,  29,31,33,35,37,39,41,43,45,47,49,51,53,55 |
| Host OS | 0,1,28,29 (first cores from each NUMA) |
| vRouter | Used for scenarios 2 + 2HT: 2,4,30,32  Reserved: 2,4,6,8,30,32,34,36 |

## Hypervisor configuration description

|  |
| --- |
| OS version  cat /etc/os-release  NAME="Red Hat Enterprise Linux Server"  VERSION="7.7 (Maipo)"  Kernel isolcpu and huge pages  # cat /proc/cmdline  BOOT\_IMAGE=/boot/vmlinuz-3.10.0-1062.12.1.el7.x86\_64 root=UUID=228c59ea-82f0-4ee8-9d03-5620e5f0fafb ro console=tty0 console=ttyS0,115200n8 crashkernel=auto rhgb quiet **isolcpus=2-27,30-55** iommu=pt intel\_iommu=on **default\_hugepagesz=1GB hugepagesz=1G hugepages=128 hugepagesz=2M hugepages=8192** skew\_tick=1 nohz=on nohz\_full=2-27,30-55 rcu\_nocbs=2-27,30-55 tuned.non\_isolcpus=30000003 intel\_pstate=disable nosoftlockup  Nova vcpu pinning  grep ^vcpu /var/lib/config-data/puppet-generated/nova\_libvirt/etc/nova/nova.conf  vcpu\_pin\_set=**10,12,14,16,18,20,22,24,26,38,40,42,44,46,48,50,52,54**  Tuned cpu-partitioning profile configuration  grep ^isolated\_cores /etc/tuned/cpu-partitioning-variables.conf  isolated\_cores=**2-27,30-55**  vRouter forwarding cores allocation  # grep ^CPU /etc/sysconfig/network-scripts/ifcfg-vhost0  CPU\_LIST=**2,4,30,32** |

# ***Test setup***



The test setup consists of three compute nodes with the above mentioned hardware specification. The device under test (DUT) is compute-B in the above setup where the vRouter is running. The vRouter is hosting a PROX VM which just loops traffic on the same interface. The traffic generator is also a PROX VM hosted by vRouter in compute-A. There is a third VM called Control traffic VM which is running in compute-C. The purpose of this VM is to control the traffic generator (start, stop, pause etc.) and also to collect statistics. In order to saturate the DUT, the traffic generator VM and compute are allocated much more resources than the DUT. The vRouter itself is running on a bonded interface in 802.3ad LACP mode.

By default on Prox generate and Prox looping VMs the multiqueue is enabled with a number of queues appropriate to vRouter queues. Enabling the multiqueue means enabling it on the openstack and VM (Prox) level.

# 

# ***About PROX and Rapid***

Prox is an OPNFV project application built on top of DPDK which allows to create flexible software architectures through small and readable configuration files. One of the ways PROX can be used is as a traffic generator for performance investigations. Rapid is a wrapper on PROX to simplify the configuration of PROX even further.

Rapid github repository:

<https://github.com/opnfv/samplevnf/blob/master/VNFs/DPPD-PROX/helper-scripts/rapid/README>

Heat templates to deploy Rapid

<https://github.com/damjanek/dpdk-prox-contrail>

## Prox VMs tuning description (compute-B)

The Prox VMs are also tuned using kernel isolcpu and tuned profile and have enabled multiqueue

|  |
| --- |
| OS version  # cat /etc/os-release  NAME="CentOS Linux"  VERSION="7 (Core)"  Kernel isolcpu  # cat /proc/cmdline  BOOT\_IMAGE=/boot/vmlinuz-3.10.0-1062.12.1.el7.x86\_64 root=UUID=3ef2b806-efd7-4eef-aaa2-2584909365ff ro console=tty0 console=ttyS0,115200n8 crashkernel=auto console=ttyS0,115200 LANG=en\_US.UTF-8 **isolcpus=1-9** intel\_pstate=disable nosoftlockup skew\_tick=1 nohz=on nohz\_full=1-9 rcu\_nocbs=1-9  Tuned realtime-virtual-guest-variables profile configuration  grep ^isolated\_cores /etc/tuned/realtime-virtual-guest-variables.conf  isolated\_cores=1-9  Multiqueue enabled in Image  openstack image show rapidVM-1908 -c properties -f value  direct\_url='rbd://27ef521a-743a-11ea-b2f9-5254001643f5/images/56aac0c7-dc10-4756-af2d-c0f115735561/snap', **hw\_vif\_multiqueue\_enabled='true'**, locations='[{u'url': u'rbd://27ef521a-743a-11ea-b2f9-5254001643f5/images/56aac0c7-dc10-4756-af2d-c0f115735561/snap', u'metadata': {}}]'  Multiqueue enabled in Prox (CPU allocation for looping VM compute-B)  **Core 2: RX port 0 (queue 0) ==> TX port 0 (queue 0)**  **Core 3: RX port 0 (queue 1) ==> TX port 0 (queue 1)**  **Core 4: RX port 0 (queue 2) ==> TX port 0 (queue 2)**  **Core 5: RX port 0 (queue 3) ==> TX port 0 (queue 3)**  Core 6: RX port 0 (queue 4) ==> TX port 0 (queue 4)  Core 7: RX port 0 (queue 5) ==> TX port 0 (queue 5)  Core 8: RX port 0 (queue 6) ==> TX port 0 (queue 6)  Core 9: RX port 0 (queue 7) ==> TX port 0 (queue 7)  VM flavor properties  openstack flavor show ii-stack3-swap -c properties -f value  **hw:cpu\_policy='dedicated', hw:mem\_page\_size='large'** |

# 

# 

# ***Test case 1:*** Number of p***ackets per*** s***econd (***Mpps***) and*** Throughput (Gbps) ***per core on Intel 82599 Niantic Ethernet Adapters***

## Configuration of vRouter on compute-B (DUT):

|  |  |
| --- | --- |
| Operating System | RedHat 7.7 |
| Number of Forwarding Cores | 2 (Physical) + 2 (Siblings) |
| Drop Rate | 0.001% |
| Bond | 2 x 10G (802.3ad LACP) with L3/L4 hash |
| Hugepages | 1G |
| NUMA allocation | vRouter, NIC and VM on the same socket (NUMA 0) |

## CPU allocation (Full-CPU partitioning scheme) on compute-B:

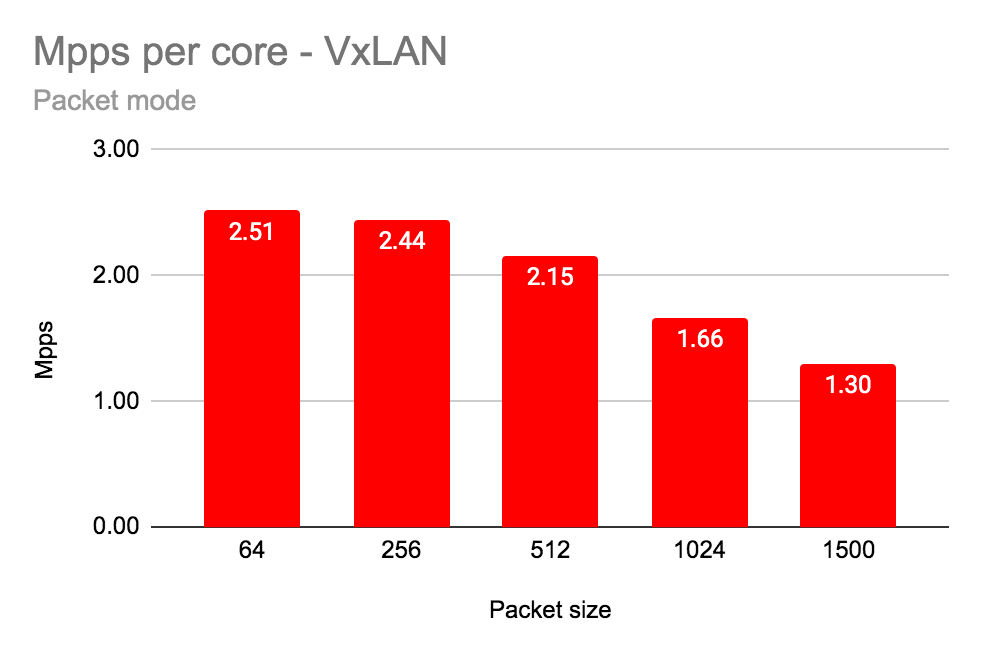
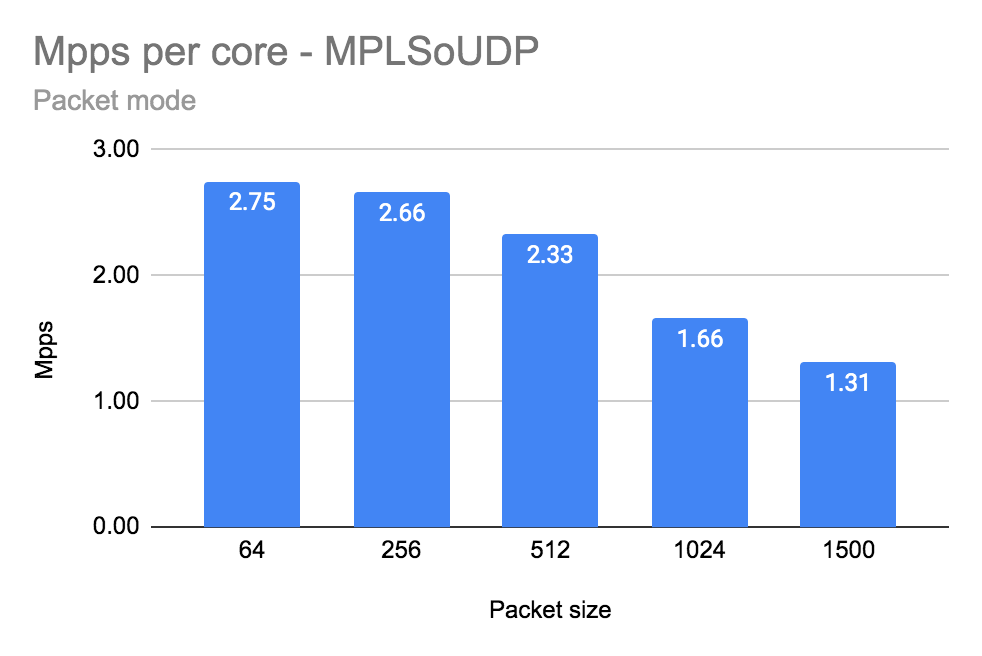
|  |  |
| --- | --- |
| vRouter forwarding cores | 2,4,30,32 |
| Service cores | 0,1,28,29 |
| Control cores | 0,1,28,29 |
| Nova (VM) cores [VM core: HV core] | 0: 18,1: 46, 2: 40, 3: 12, 4: 48, 5: 20, 6: 24, 7: 52, 8: 10, 9: 38 |
| Host OS cores | 0,1,28,29 |
| Kernel isolcpus and tuned isolcpu list | 2-27,30-55 |

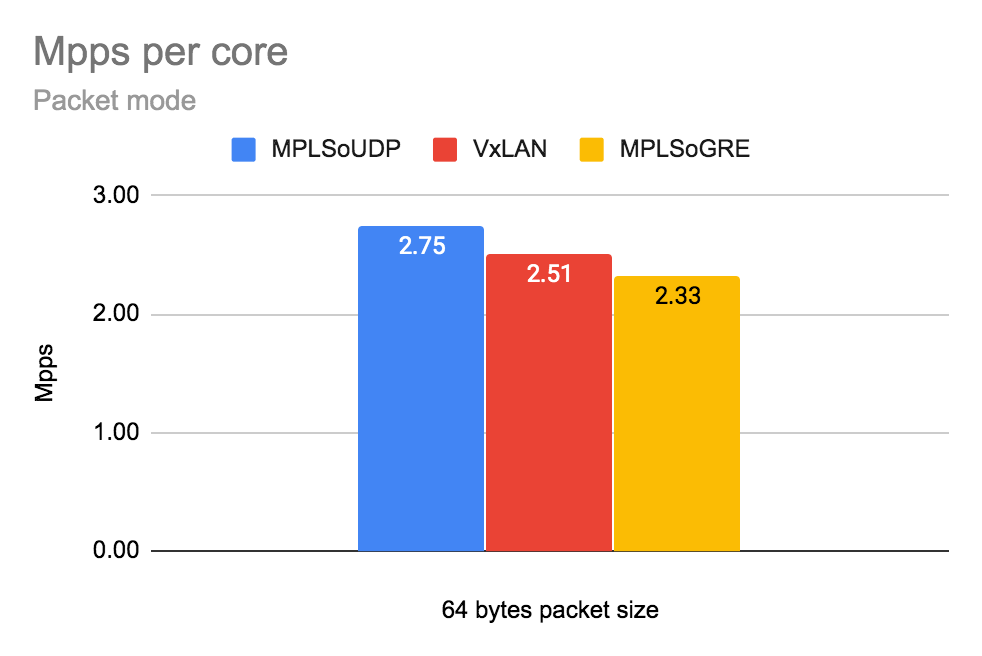
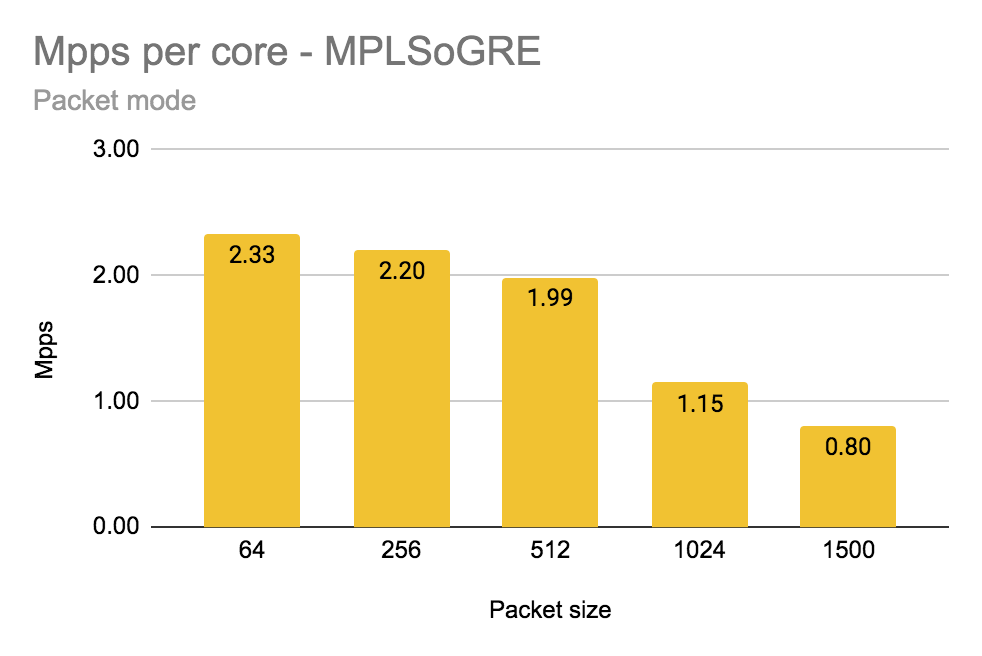
## 

## Traffic pattern

|  |  |
| --- | --- |
| Number of flows | 16535 |
| Encapsulation | MPLSoUDP, MPLSoGRE, VxLAN |
| Packet size | 64B, 256B, 512B, 1024B, 1500B |

## P*acket mode (Mpps):*



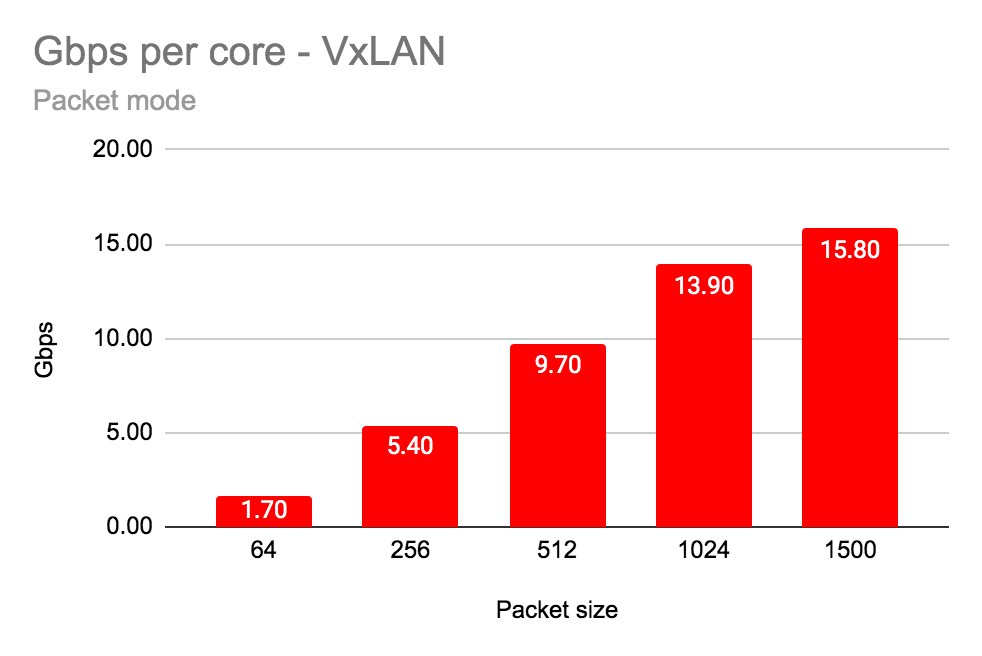
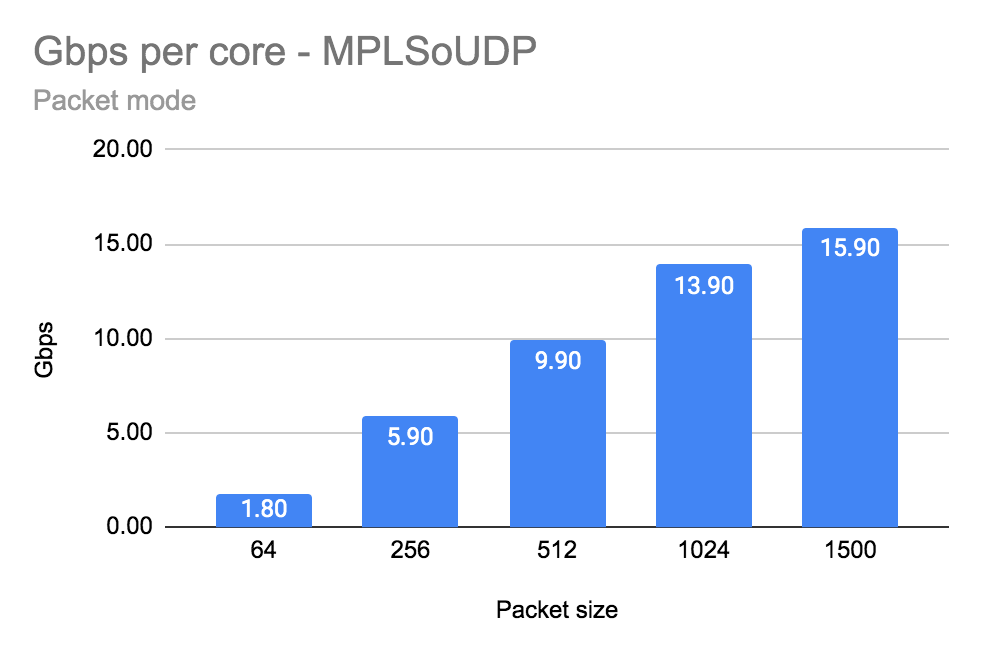


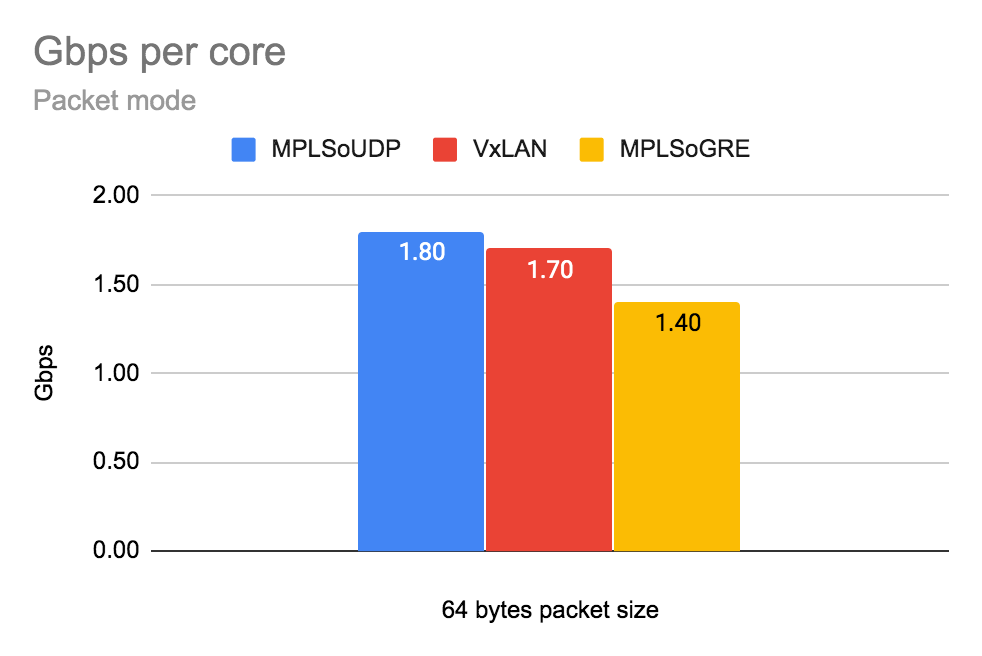
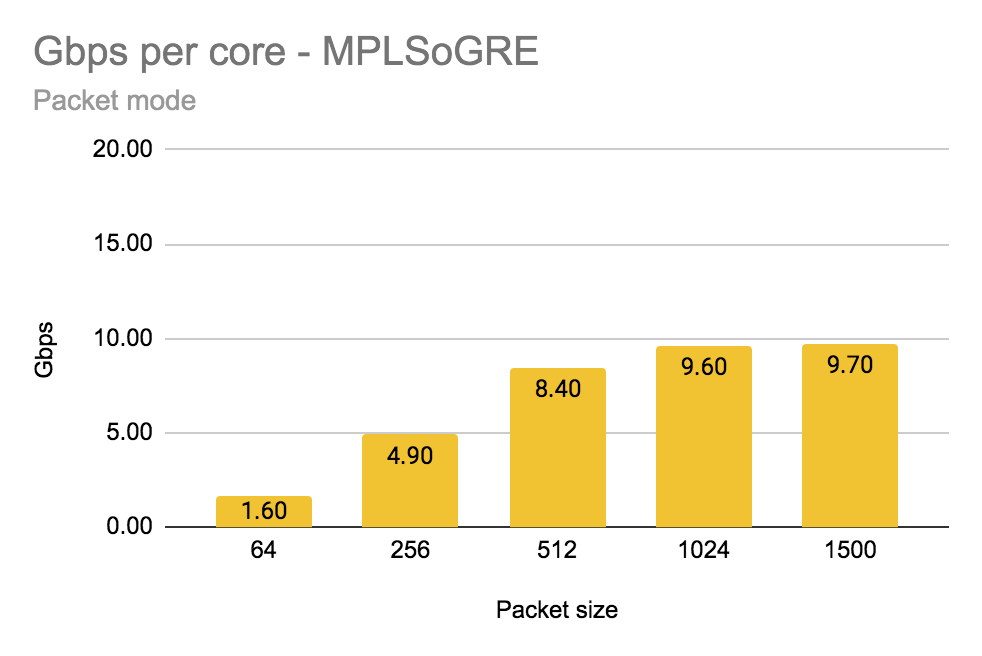
Number of packets per second (Mpps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Encapsulation | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| MPLSoUDP | 2.75 | 2.66 | 2.33 | 1.66 | 1.31 |
| MPLSoGRE | 2.33 | 2.20 | 1.99 | 1.15\* | 0.80\* |
| VxLAN | 2.51 | 2.44 | 2.15 | 1.66 | 1.30 |

\* not enough entropy on LACP causes only one 10G link utilisation

## Packet mode (Gbps):



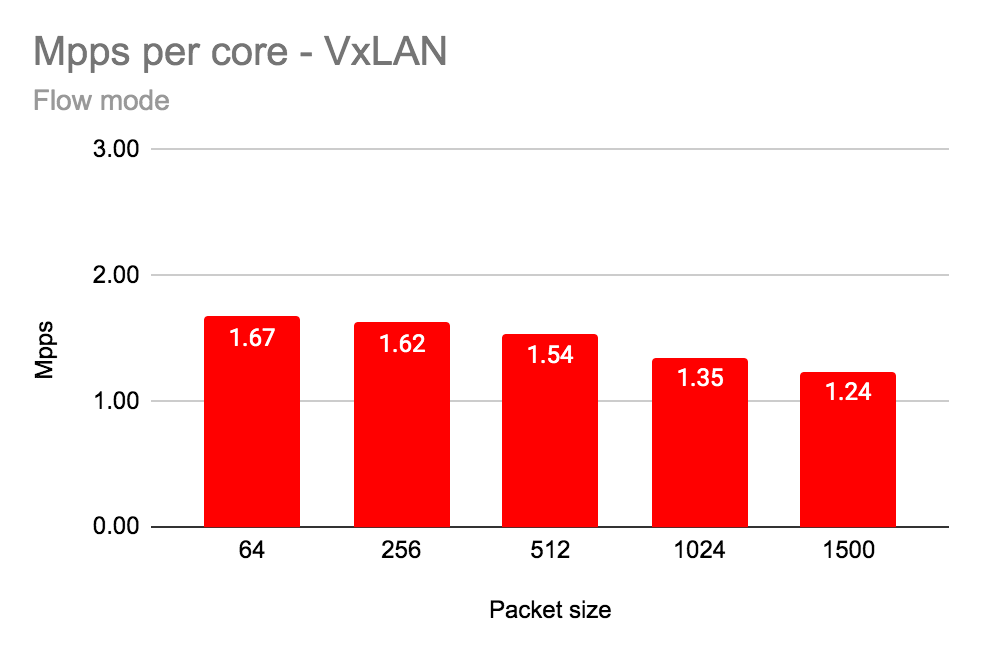
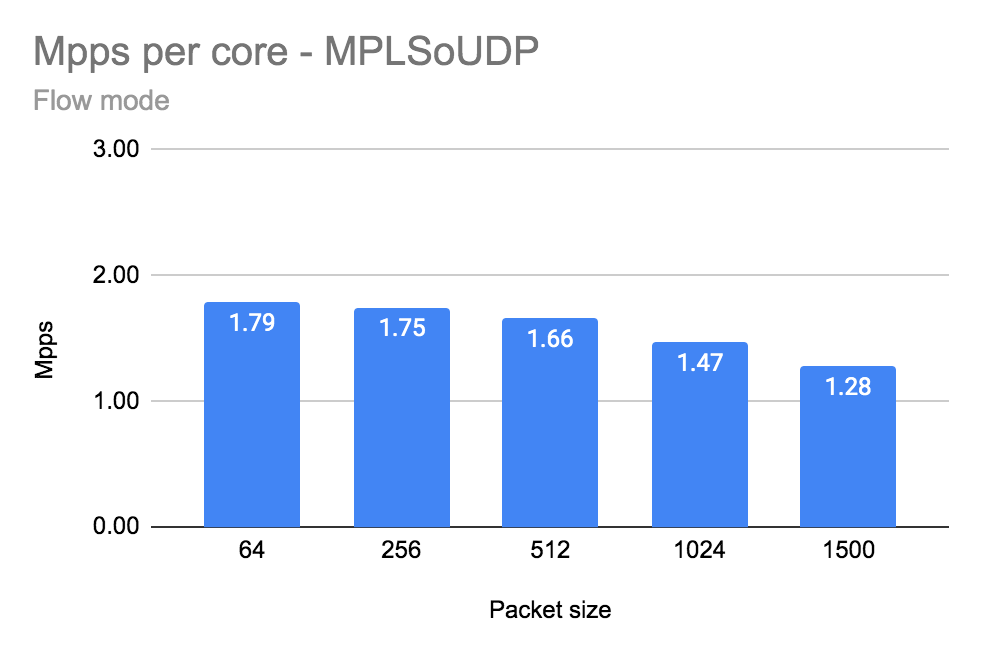


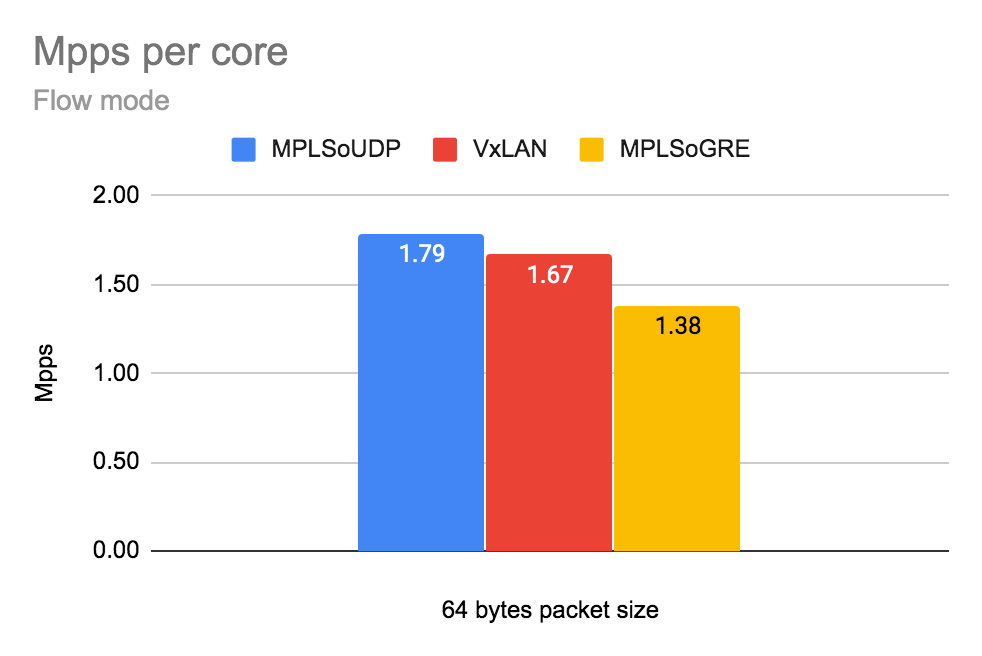
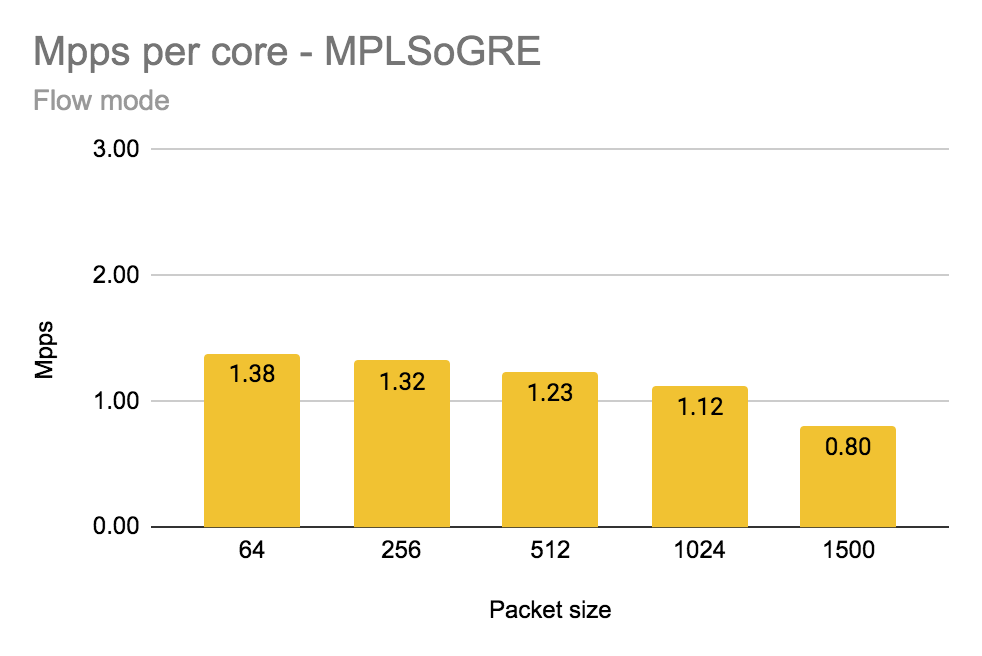
Number of gigabits per second (Gbps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Encapsulation | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| MPLSoUDP | 1.80 | 5.90 | 9.90 | 13.90 | 15.90 |
| MPLSoGRE | 1.60 | 4.90 | 8.40 | 9.60\* | 9.70\* |
| VxLAN | 1.70 | 5.40 | 9.70 | 13.90 | 15.80 |

\* not enough entropy on LACP causes only one 10G link utilisation

## Flow mode (Mpps):





Number of packets per second (mpps) per physical core (with siblings).

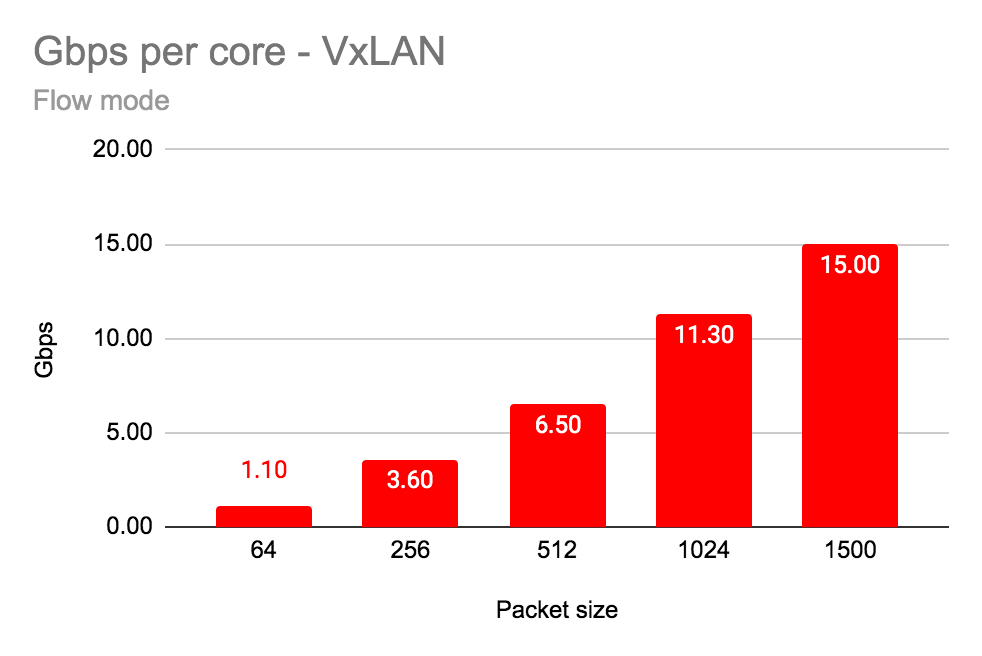
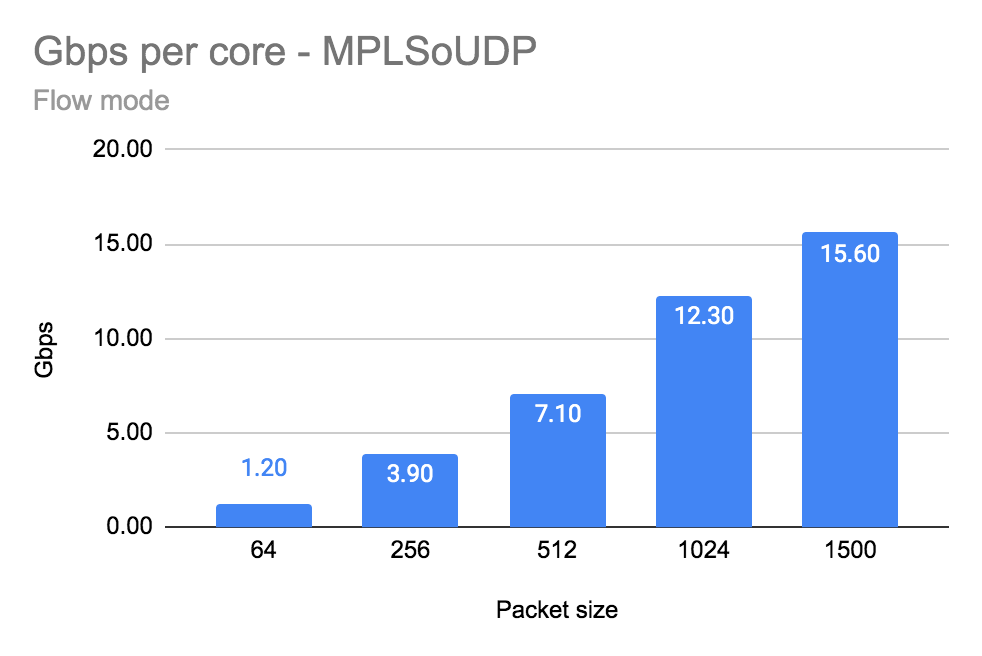
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Encapsulation | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| MPLSoUDP | 1.79 | 1.75 | 1.66 | 1.47 | 1.28 |
| MPLSoGRE | 1.38 | 1.32 | 1.23 | 1.12\* | 0.80\* |
| VxLAN | 1.67 | 1.62 | 1.54 | 1.35 | 1.24 |

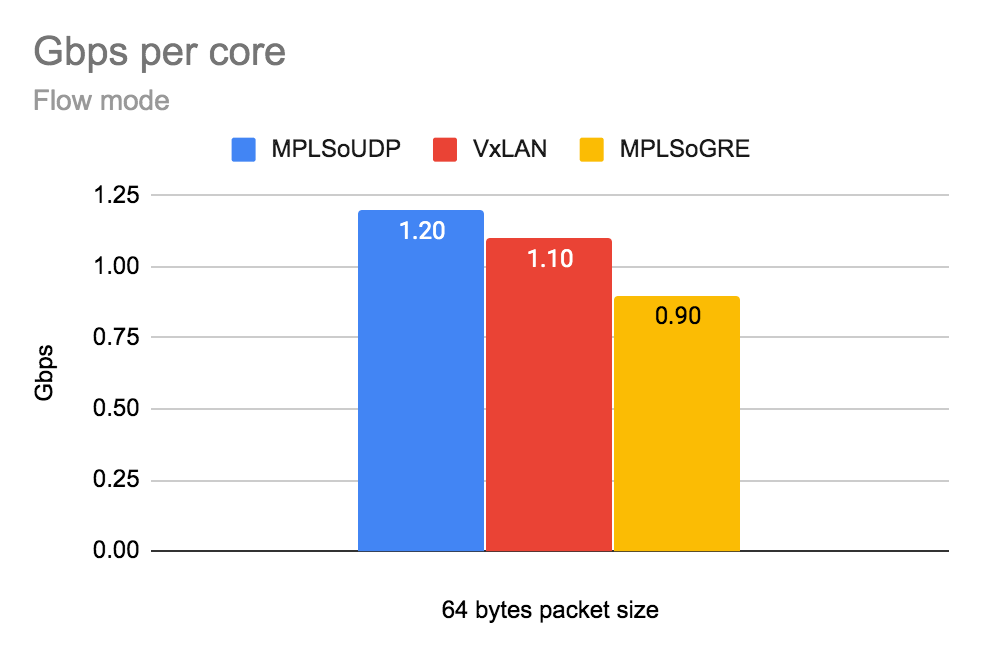
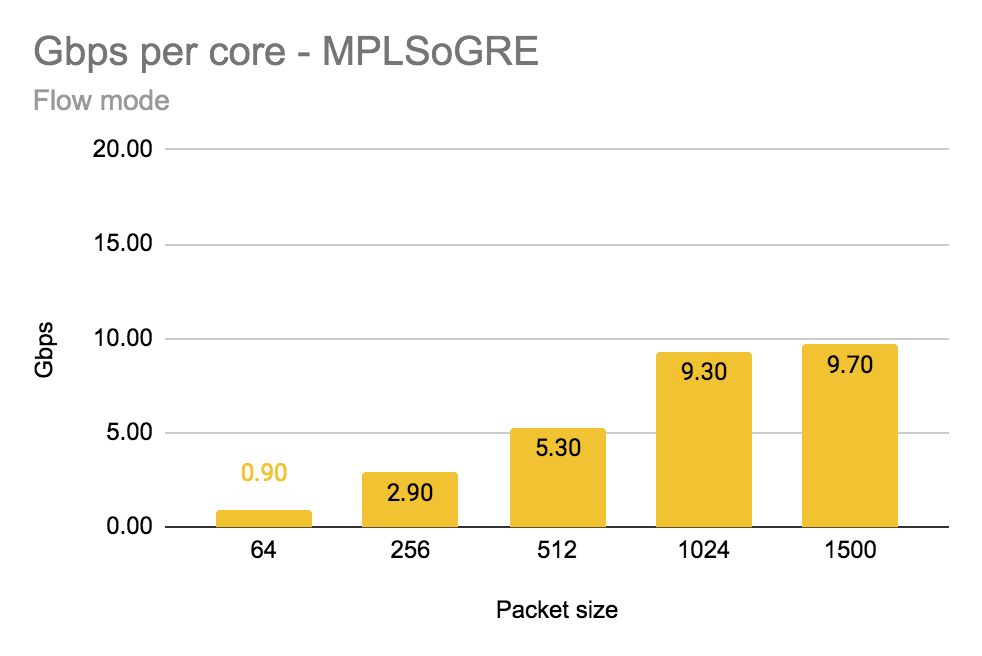
\* not enough entropy on LACP causes only one 10G link utilisation

## 

## 

## Flow mode (Gbps):





Number of gigabits per second (Gbps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Encapsulation | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| MPLSoUDP | 1.20 | 3.90 | 7.10 | 12.30 | 15.60 |
| MPLSoGRE | 0.90 | 2.90 | 5.30 | 9.30\* | 9.70\* |
| VxLAN | 1.10 | 3.60 | 6.50 | 11.30 | 15.00 |

\* not enough entropy on LACP causes only one 10G link utilisation

## 

## Summary

The fastest encapsulation is MPLSoUDP in both cases - packet and flow mode. Due to various source ports of encapsulated MPLSoUDP and VxLAN traffic, both 10G links were used by the LACP hash algorithm. In case of MPLSoGRE not enough entropy causes only one LACP link saturation and in consequence poor numbers for packet sizes >1024 bytes (reached one 10G interface limit).

MPLSoUDP encapsulation latency is high, reducing a speed (pps/per core) around 5% cause reduction of latency around 300-400%. Detailed numbers provided in the case 2a.

## 

## 82599 Niantic configuration

The NIC specific configuration parameters

|  |
| --- |
| lshw -c network -businfo| grep 82599ES  pci@0000:5e:00.0 network 82599ES 10-Gigabit SFI/SFP+ Network Connection  pci@0000:5e:00.1 network 82599ES 10-Gigabit SFI/SFP+ Network Connection  cat /etc/sysconfig/network-scripts/ifcfg-vhost0  [...]  BIND\_INT=0000:5e:00.0,0000:5e:00.1  DRIVER=uio\_pci\_generic  BOND\_MODE=4  BOND\_POLICY=layer3+4  CPU\_LIST=2,4,30,32  SERVICE\_CORE\_MASK=0,1,28,29  DPDK\_CTRL\_THREAD\_MASK=0,1,28,29  sed -e "s/\x00/ /g" /proc/$(pidof contrail-vrouter-dpdk)/cmdline  /usr/bin/contrail-vrouter-dpdk --no-daemon --vr\_flow\_entries=2500000 --vr\_mempool\_sz 262144 --vr\_dpdk\_rx\_ring\_sz 2048 --vr\_dpdk\_tx\_ring\_sz 2048 --yield\_option 0 --service\_core\_mask 0x30000003 --dpdk\_ctrl\_thread\_mask 0x30000003 --socket-mem 1024 1024 --vdev eth\_bond\_bond1,mode=4,xmit\_policy=l34,socket\_id=0,mac=90:e2:ba:e3:5d:04,lacp\_rate=1,slave=0000:5e:00.0,slave=0000:5e:00.1 |

## Detailed data

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **vRouter Mode** | **Data** | **Encapsulation** | **Packet size [bytes]** | | | | |
|  | **64** | **256** | **512** | **1024** | **1500** |
| Packet Mode | Total Throughput Gbps [TX+RX] | MPLSoUDP | 3.60 | 11.80 | 19.80 | 27.80 | 31.80 |
| MPLSoGRE | 3.20 | 9.80 | 16.80 | 19.20 | 19.40 |
| VxLAN | 3.40 | 10.80 | 19.40 | 27.80 | 31.60 |
| Throughput Gbps/core | MPLSoUDP | 1.80 | 5.90 | 9.90 | 13.90 | 15.90 |
| MPLSoGRE | 1.60 | 4.90 | 8.40 | 9.60 | 9.70 |
| VxLAN | 1.70 | 5.40 | 9.70 | 13.90 | 15.80 |
| Total Mpps [TX+RX] | MPLSoUDP | 5.49 | 5.32 | 4.65 | 3.33 | 2.62 |
| MPLSoGRE | 4.65 | 4.40 | 3.97 | 2.30 | 1.60 |
| VxLAN | 5.03 | 4.87 | 4.30 | 3.33 | 2.60 |
| Mpps/core | MPLSoUDP | 2.75 | 2.66 | 2.33 | 1.66 | 1.31 |
| MPLSoGRE | 2.33 | 2.20 | 1.99 | 1.15 | 0.80 |
| VxLAN | 2.51 | 2.44 | 2.15 | 1.66 | 1.30 |
| Avg. Latency | MPLSoUDP | 862 | 628 | 191 | 142 | 130 |
| MPLSoGRE | 968 | 850 | 531 | 264 | 253 |
| VxLAN | 816 | 837 | 184 | 170 | 142 |
| Flow Mode | Total Throughput Gbps [TX+RX] | MPLSoUDP | 2.40 | 7.80 | 14.20 | 24.60 | 31.20 |
| MPLSoGRE | 1.80 | 5.80 | 10.60 | 18.60 | 19.40 |
| VxLAN | 2.20 | 7.20 | 13.00 | 22.60 | 30.00 |
| Throughput Gbps/core | MPLSoUDP | 1.20 | 3.90 | 7.10 | 12.30 | 15.60 |
| MPLSoGRE | 0.90 | 2.90 | 5.30 | 9.30 | 9.70 |
| VxLAN | 1.10 | 3.60 | 6.50 | 11.30 | 15.00 |
| Total Mpps [TX+RX] | MPLSoUDP | 3.57 | 3.49 | 3.31 | 2.95 | 2.56 |
| MPLSoGRE | 2.76 | 2.64 | 2.47 | 2.24 | 1.60 |
| VxLAN | 3.34 | 3.25 | 3.07 | 2.70 | 2.47 |
| Mpps/core | MPLSoUDP | 1.79 | 1.75 | 1.66 | 1.47 | 1.28 |
| MPLSoGRE | 1.38 | 1.32 | 1.23 | 1.12 | 0.80 |
| VxLAN | 1.67 | 1.62 | 1.54 | 1.35 | 1.24 |
| Avg. Latency | MPLSoUDP | 271 | 321 | 290 | 257 | 233 |
| MPLSoGRE | 787 | 673 | 191 | 158 | 120 |
| VxLAN | 212 | 213 | 219 | 236 | 259 |

# ***Test case 2:*** Number of packets per second (Mpps) and Throughput (Gbps) per core on ***Intel X710 Fortville Ethernet Adapters***

## Configuration of vRouter on compute-B (DUT):

|  |  |
| --- | --- |
| Operating System | RedHat 7.7 |
| Number of Forwarding Cores | 2 (Physical) + 2 (Siblings) |
| Drop Rate | 0.001% |
| Bond | 2 x 10G (802.3ad LACP) with L3/L4 hash |
| Hugepages | 1G |
| NUMA allocation | vRouter, NIC and VM on the same socket (NUMA 0) |

## CPU allocation (Full-CPU partitioning scheme) on compute-B:

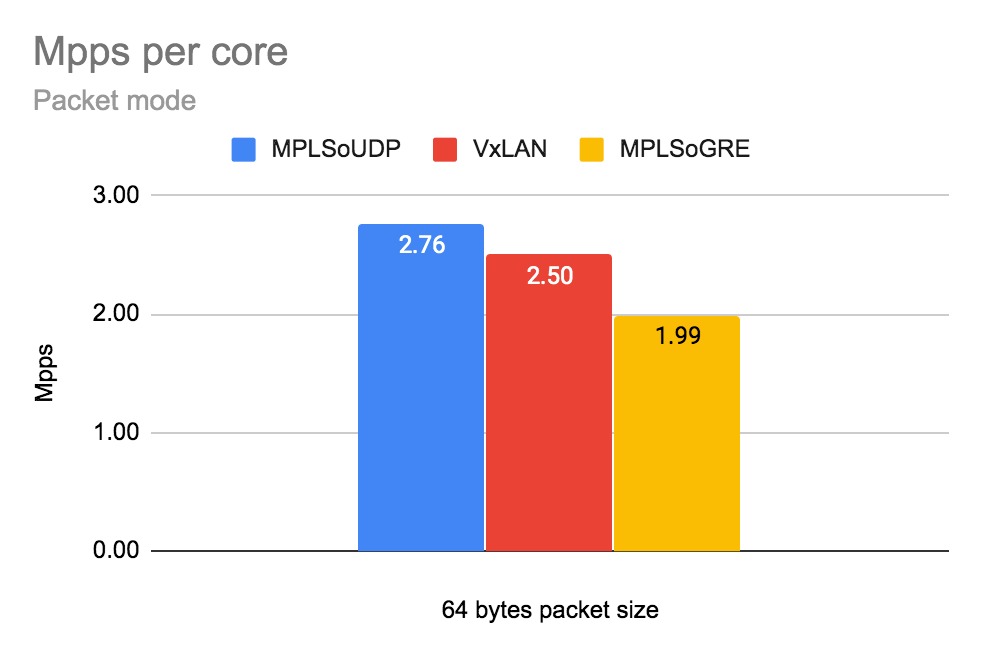
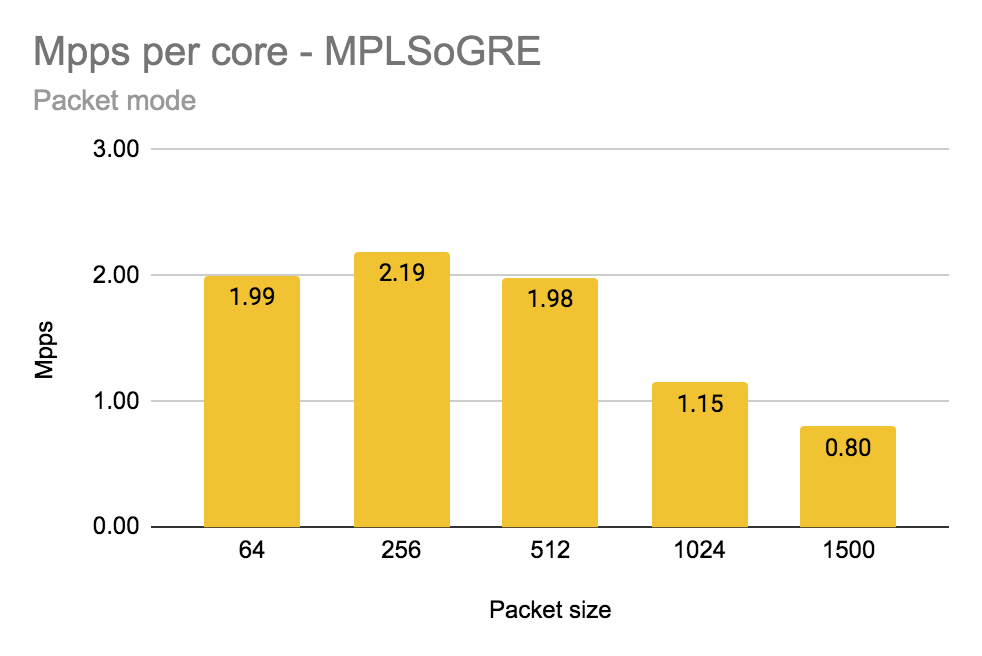
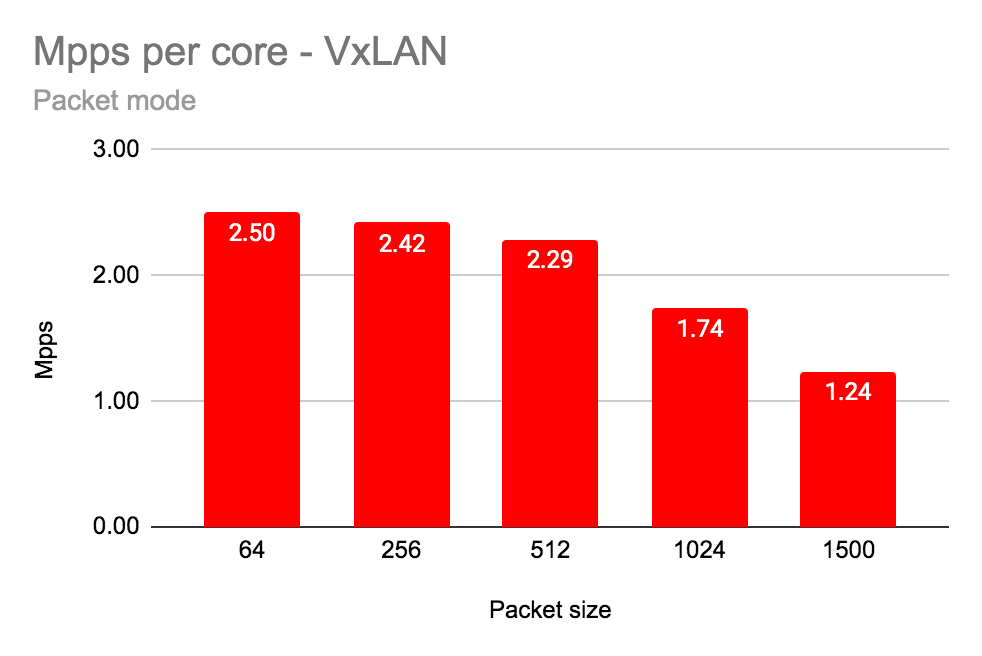
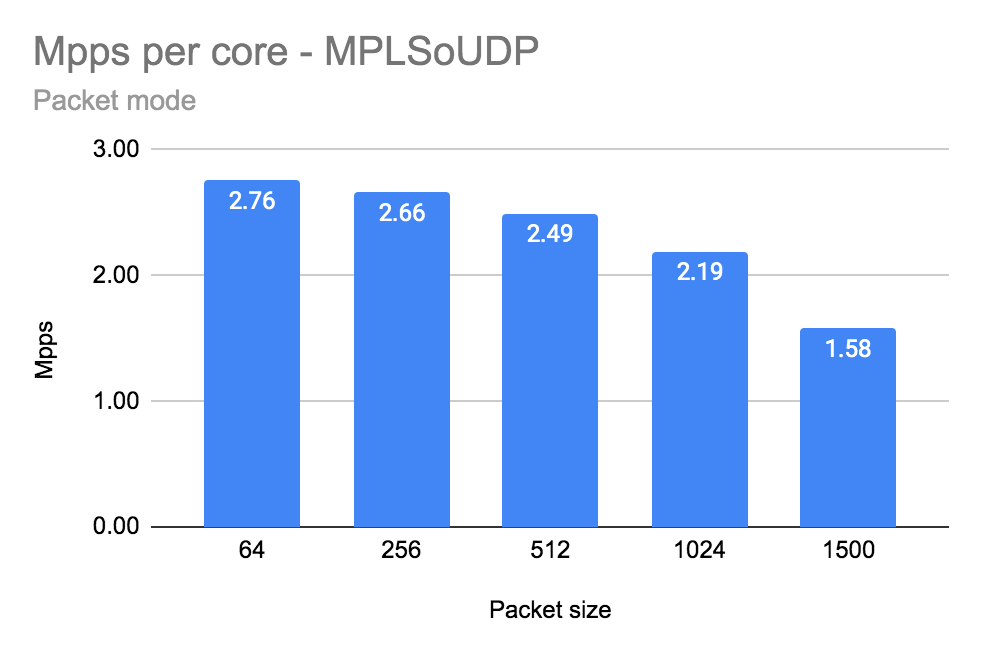
|  |  |
| --- | --- |
| vRouter forwarding cores | 2,4,30,32 |
| Service cores | 0,1,28,29 |
| Control cores | 0,1,28,29 |
| Nova (VM) cores [VM core: HV core] | 0: 18,1: 46, 2: 40, 3: 12, 4: 48, 5: 20, 6: 24, 7: 52, 8: 10, 9: 38 |
| Host OS cores | 0,1,28,29 |
| Kernel isolcpus and tuned isolcpu list | 2-27,30-55 |

## 

## Traffic pattern

|  |  |
| --- | --- |
| Number of flows | 16535 |
| Encapsulation | MPLSoUDP, MPLSoGRE, VxLAN |
| Packet sizes | 64B, 256B, 512B, 1024B, 1500B |

## Packet mode (Mpps):

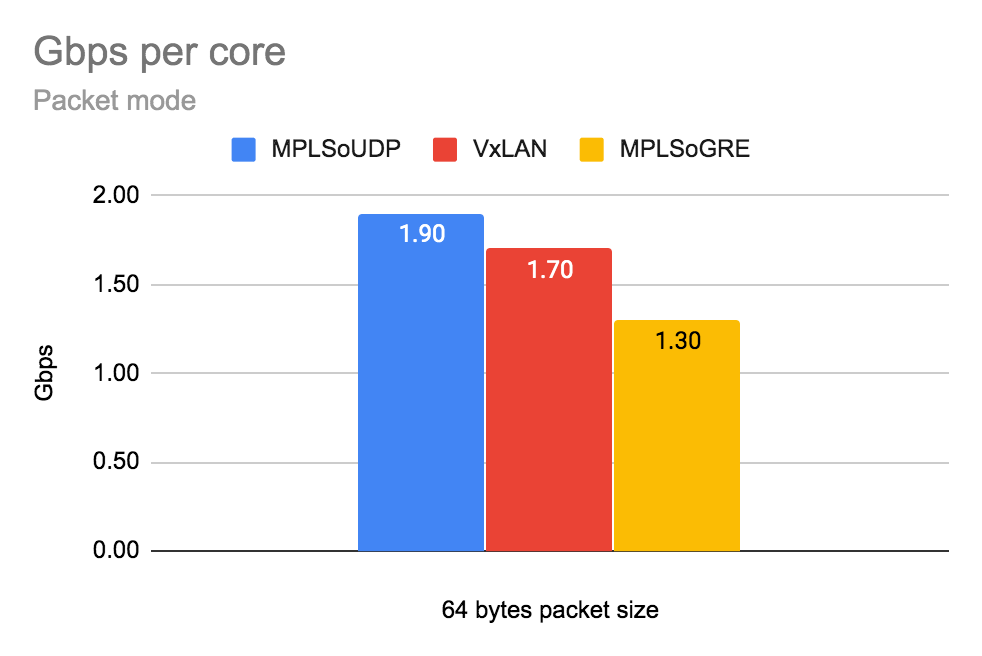
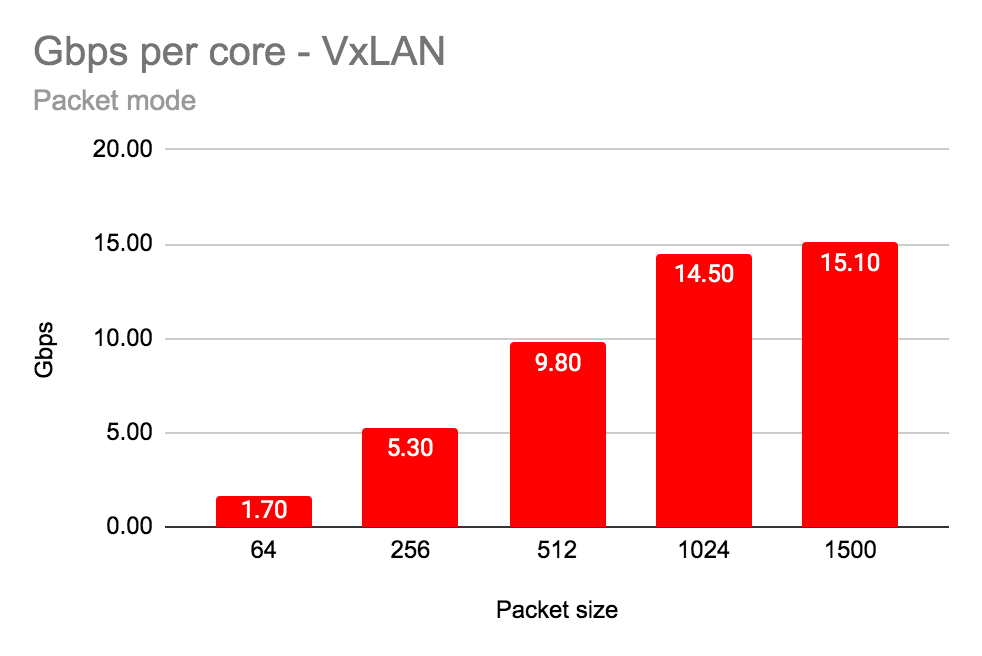
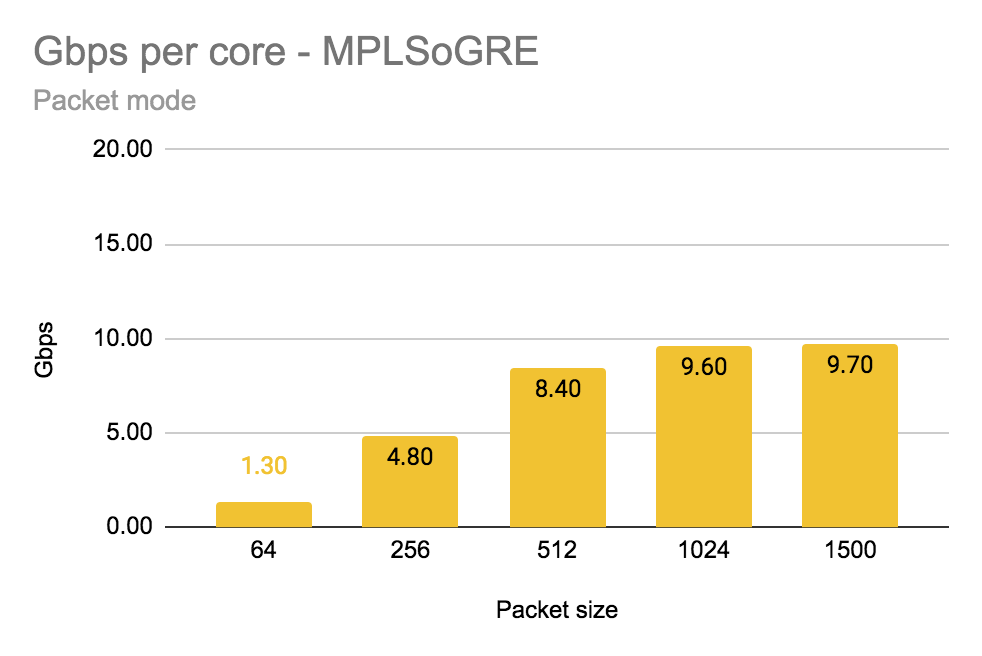
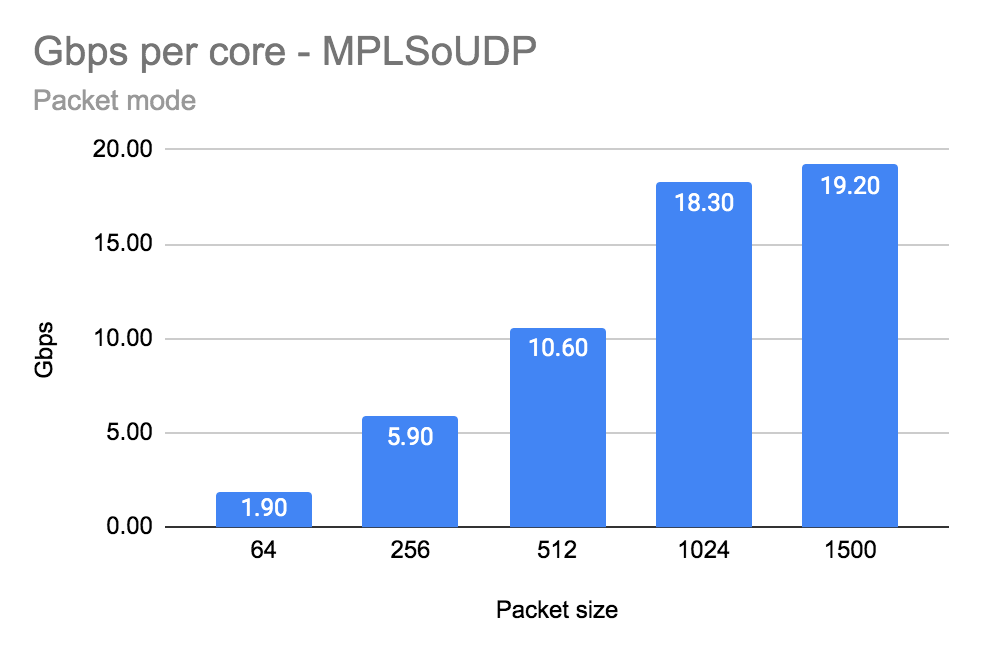


Number of packets per second (Mpps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Encapsulation | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| MPLSoUDP | 2.76 | 2.66 | 2.49 | 2.19 | 1.58 |
| MPLSoGRE | 1.99 | 2.19 | 1.98 | 1.15 | 0.80 |
| VxLAN | 2.50 | 2.42 | 2.29 | 1.74 | 1.24 |

## 

## Packet mode (Gbps):

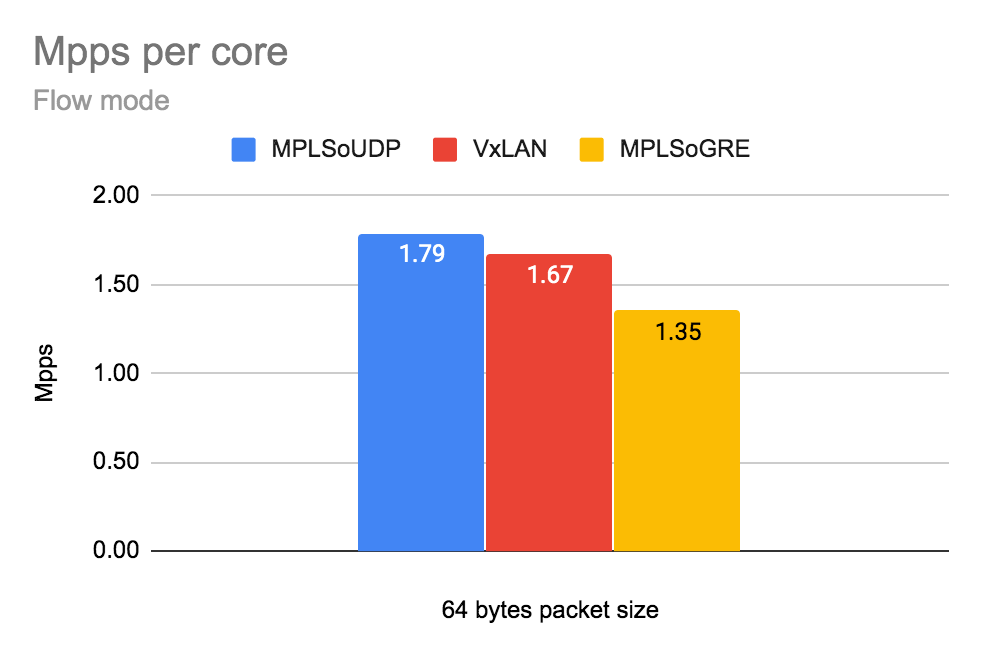
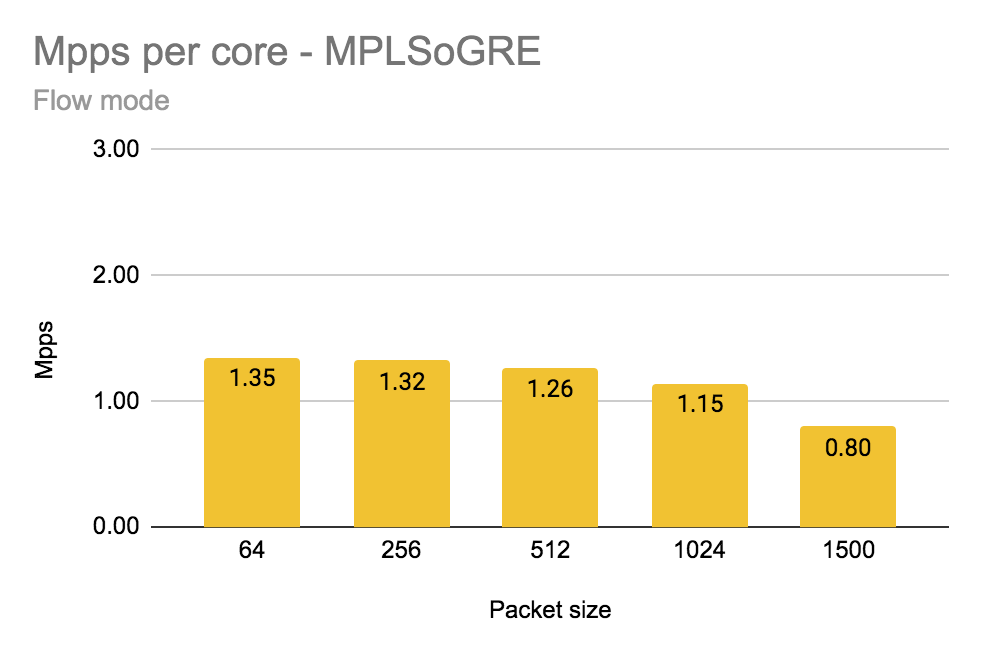
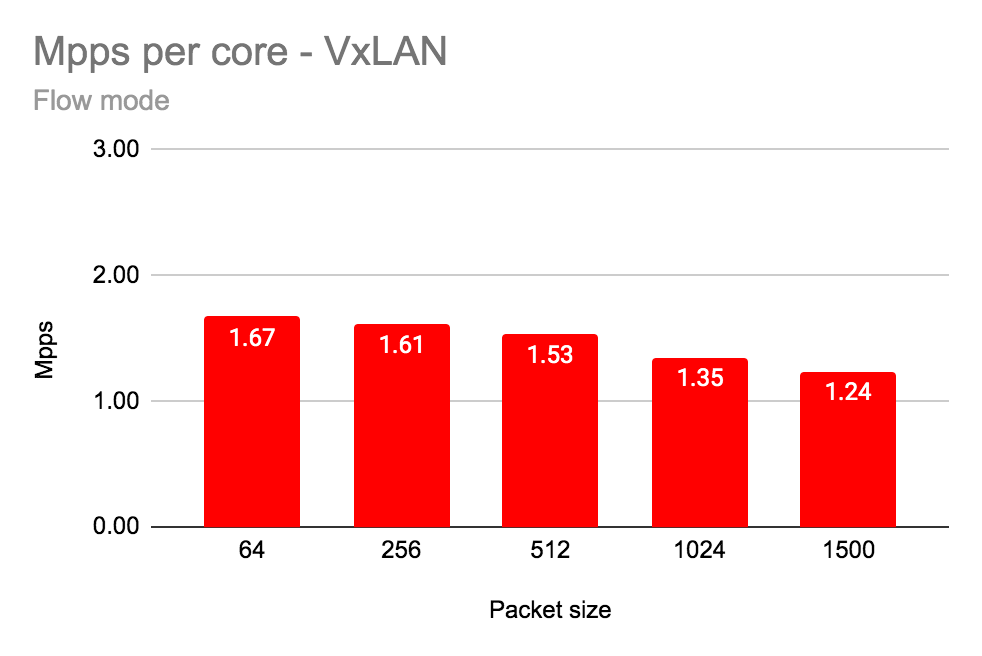
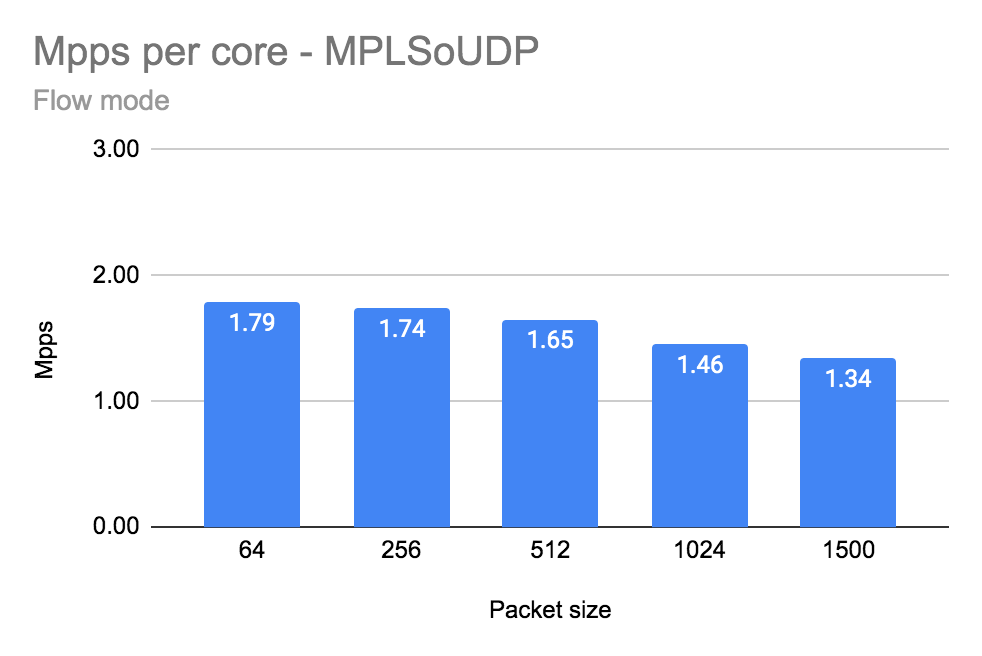


Number of gigabits per second (Gbps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Encapsulation | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| MPLSoUDP | 1.90 | 5.90 | 10.60 | 18.30 | 19.20 |
| MPLSoGRE | 1.30 | 4.80 | 8.40 | 9.60\* | 9.70\* |
| VxLAN | 1.70 | 5.30 | 9.80 | 14.50 | 15.10 |

\* not enough entropy on LACP causes only one 10G link utilisation

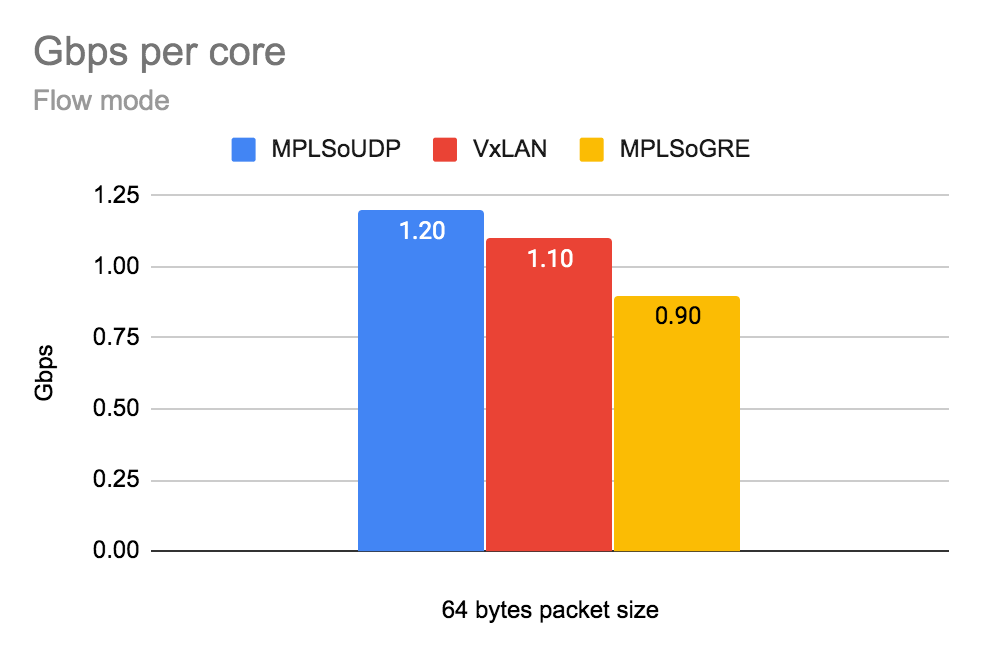
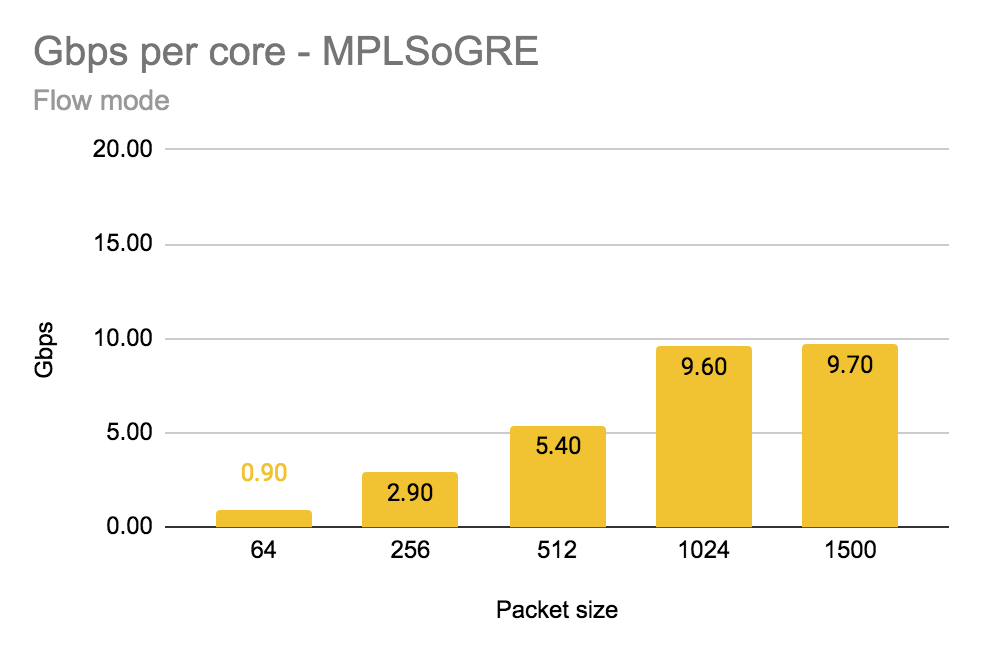
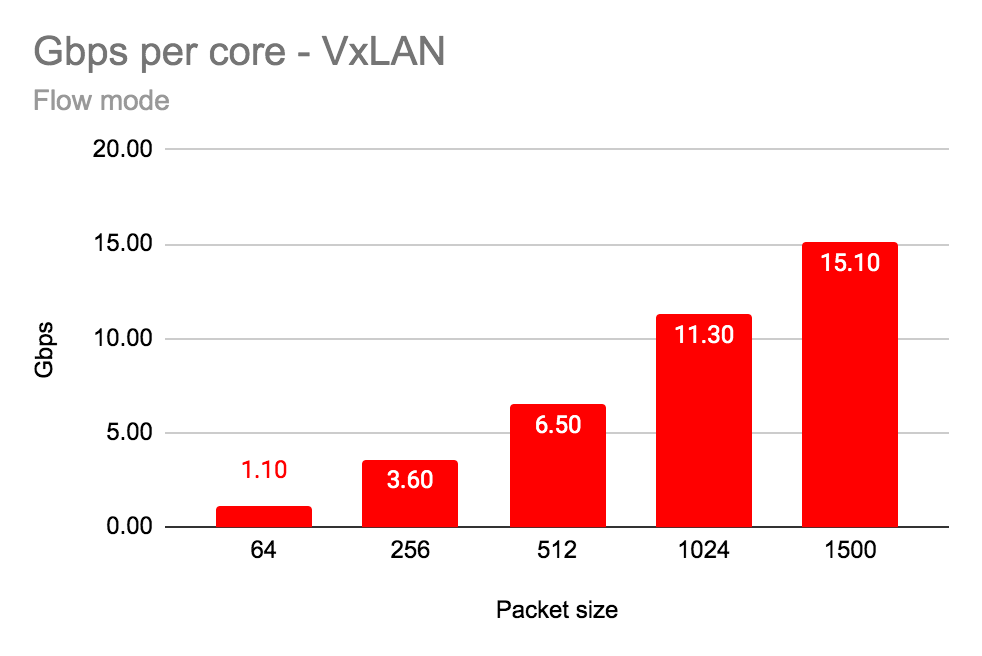
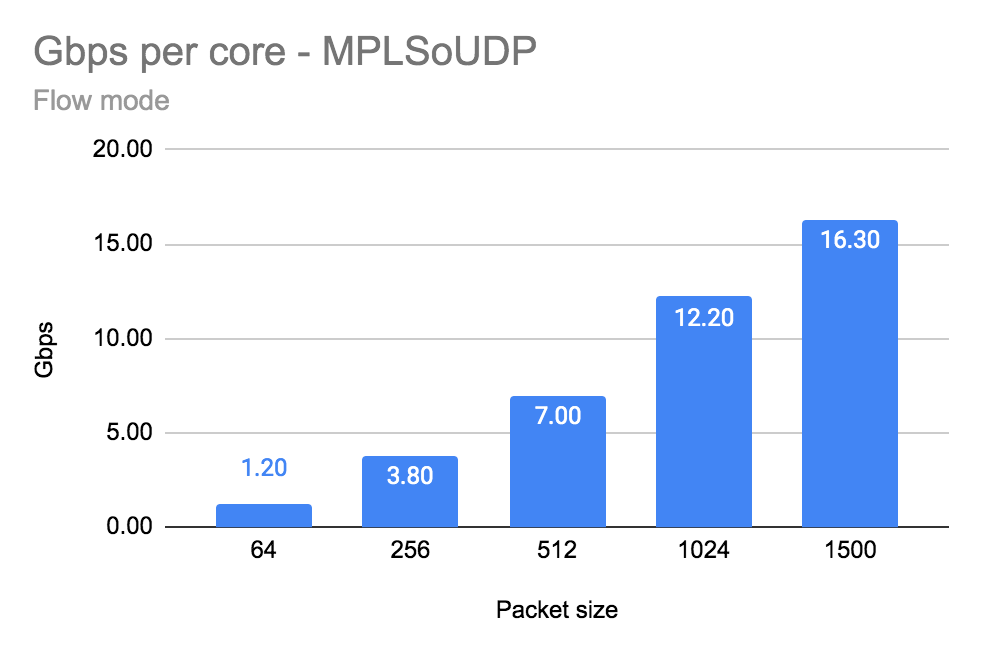
## Flow mode (Mpps):



Number of packets per second (Mpps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Encapsulation | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| MPLSoUDP | 1.79 | 1.74 | 1.65 | 1.46 | 1.34 |
| MPLSoGRE | 1.35 | 1.32 | 1.26 | 1.15 | 0.80 |
| VxLAN | 1.67 | 1.61 | 1.53 | 1.35 | 1.24 |

## Flow mode (Gbps):



Number of gigabits per second (Gbps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Encapsulation | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| MPLSoUDP | 1.20 | 3.80 | 7.00 | 12.20 | 16.30 |
| MPLSoGRE | 0.90 | 2.90 | 5.40 | 9.60\* | 9.70\* |
| VxLAN | 1.10 | 3.60 | 6.50 | 11.30 | 15.10 |

\* not enough entropy on LACP causes only one 10G link utilisation

## Summary

The fastest encapsulation is MPLSoUDP in both cases - packet and flow mode. Due to various source ports of encapsulated MPLSoUDP and VxLAN traffic, both 10G links were used by the LACP hash algorithm. In case of MPLSoGRE not enough entropy causes only one LACP link saturation and in consequence poor numbers for packet sizes >1024 bytes (reached one 10G interface limit).

MPLSoUDP encapsulation latency is high, reducing a speed (pps/per core) around 5% cause reduction of latency around 300-400%. Detailed numbers provided in the case 2a.

On a production with bursty traffic it is recommended to configure larger buffer descriptors size (RX/TX) towards NIC cards using parameters: *dpdk\_txd\_sz and dpdk\_rxd\_sz and adjust memory pool size* vr\_mempool\_sz (default value is 128).

The formula to calculate vr\_mempool\_sz:

**mempool size** = *2 \* (number\_of\_RX\_descriptors + number\_of\_TX\_descriptors) \* number\_of\_vrouter\_cores \* number\_of\_ports\_in\_dpdk\_bond*

## 

## Intel x710 Fortville configuration

The NIC specific configuration parameters

|  |
| --- |
| cat /etc/sysconfig/network-scripts/ifcfg-vhost0  [...]  BIND\_INT=0000:19:00.0,0000:19:00.1  DRIVER=vfio-pci  BOND\_MODE=4  BOND\_POLICY=layer3+4  CPU\_LIST=2,4,30,32  SERVICE\_CORE\_MASK=0,1,28,29  DPDK\_CTRL\_THREAD\_MASK=0,1,28,29  sed -e "s/\x00/ /g" /proc/$(pidof contrail-vrouter-dpdk)/cmdline  /usr/bin/contrail-vrouter-dpdk --no-daemon --vr\_flow\_entries=2500000 --vr\_mempool\_sz 262144 --dpdk\_ctrl\_thread\_mask 0 --vr\_dpdk\_rx\_ring\_sz 2048 --vr\_dpdk\_tx\_ring\_sz 2048 --yield\_option 0 --service\_core\_mask (0,1,28,29) --dpdk\_ctrl\_thread\_mask (0,1,28,29) --socket-mem 1024 1024 --vdev eth\_bond\_bond1,mode=4,xmit\_policy=l34,socket\_id=0,mac=e4:43:4b:6e:70:b0,lacp\_rate=1,slave=0000:19:00.0,slave=0000:19:00.1 |

## Detailed data

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **vRouter Mode** | **Data** | **Encapsulation** | **Frame size [bytes]** | | | | |
|  | **64** | **256** | **512** | **1024** | **1500** |
| Packet Mode | Total Throughput Gbps [TX+RX] | MPLSoUDP | 3.80 | 11.80 | 21.20 | 36.60 | 38.40 |
| MPLSoGRE | 2.60 | 9.60 | 16.80 | 19.20 | 19.40 |
| VxLAN | 3.40 | 10.60 | 19.60 | 29.00 | 30.20 |
| Throughput Gbps/core | MPLSoUDP | 1.90 | 5.90 | 10.60 | 18.30 | 19.20 |
| MPLSoGRE | 1.30 | 4.80 | 8.40 | 9.60 | 9.70 |
| VxLAN | 1.70 | 5.30 | 9.80 | 14.50 | 15.10 |
| Total Mpps [TX+RX] | MPLSoUDP | 5.52 | 5.33 | 4.99 | 4.39 | 3.16 |
| MPLSoGRE | 3.98 | 4.38 | 3.96 | 2.30 | 1.60 |
| VxLAN | 5.00 | 4.84 | 4.57 | 3.48 | 2.48 |
| Mpps/core | MPLSoUDP | 2.76 | 2.66 | 2.49 | 2.19 | 1.58 |
| MPLSoGRE | 1.99 | 2.19 | 1.98 | 1.15 | 0.80 |
| VxLAN | 2.50 | 2.42 | 2.29 | 1.74 | 1.24 |
| Avg. Latency | MPLSoUDP | 718 | 487 | 379 | 286 | 228 |
| MPLSoGRE | 85 | 619 | 322 | 100 | 93 |
| VxLAN | 831 | 578 | 461 | 184 | 134 |
| Flow Mode | Total Throughput Gbps [TX+RX] | MPLSoUDP | 2.40 | 7.60 | 14.00 | 24.40 | 32.60 |
| MPLSoGRE | 1.80 | 5.80 | 10.80 | 19.20 | 19.40 |
| VxLAN | 2.20 | 7.20 | 13.00 | 22.60 | 30.20 |
| Throughput Gbps/core | MPLSoUDP | 1.20 | 3.80 | 7.00 | 12.20 | 16.30 |
| MPLSoGRE | 0.90 | 2.90 | 5.40 | 9.60 | 9.70 |
| VxLAN | 1.10 | 3.60 | 6.50 | 11.30 | 15.10 |
| Total Mpps [TX+RX] | MPLSoUDP | 3.57 | 3.49 | 3.30 | 2.92 | 2.68 |
| MPLSoGRE | 2.70 | 2.64 | 2.53 | 2.29 | 1.60 |
| VxLAN | 3.34 | 3.22 | 3.06 | 2.69 | 2.48 |
| Mpps/core | MPLSoUDP | 1.79 | 1.74 | 1.65 | 1.46 | 1.34 |
| MPLSoGRE | 1.35 | 1.32 | 1.26 | 1.15 | 0.80 |
| VxLAN | 1.67 | 1.61 | 1.53 | 1.35 | 1.24 |
| Avg. Latency | MPLSoUDP | 323 | 292 | 256 | 218 | 242 |
| MPLSoGRE | 727 | 631 | 403 | 172 | 115 |
| VxLAN | 211 | 194 | 204 | 226 | 260 |

# Test case 2a: Number of packets per second (Mpps) and Throughput (Gbps) Intel X710 Ethernet Adapters - low latency

## Configuration of vRouter on compute-B (DUT):

|  |  |
| --- | --- |
| Operating System | RedHat 7.7 |
| Number of Forwarding Cores | 2 (Physical) + 2 (Siblings) |
| Drop Rate | 0.001% |
| Bond | 2 x 10G (802.3ad LACP) with L3/L4 hash |
| Hugepages | 1G |
| NUMA allocation | vRouter, NIC and VM on the same socket (NUMA 0) |

## CPU allocation (Full-CPU partitioning scheme) on compute-B:

|  |  |
| --- | --- |
| vRouter forwarding cores | 2,4,30,32 |
| Service cores | 0,1,28,29 |
| Control cores | 0,1,28,29 |
| Nova (VM) cores [VM core: HV core] | 0: 18,1: 46, 2: 40, 3: 12, 4: 48, 5: 20, 6: 24, 7: 52, 8: 10, 9: 38 |
| Host OS cores | 0,1,28,29 |
| Kernel isolcpus and tuned isolcpu list | 2-27,30-55 |

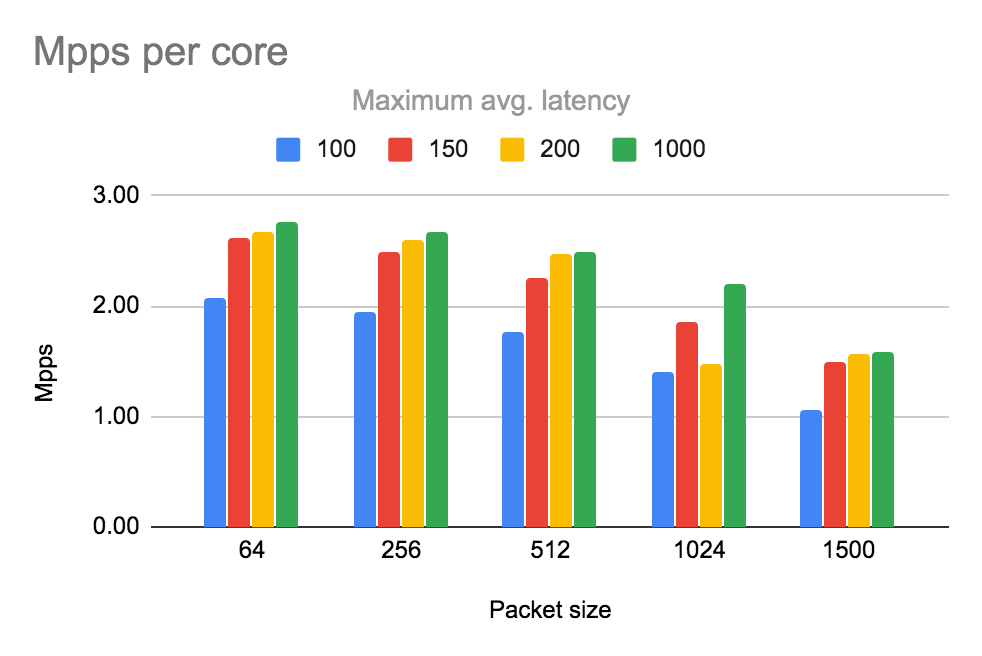
## 

## Traffic pattern

|  |  |
| --- | --- |
| Number of flows | 16535 |
| Encapsulation | MPLSoUDP |
| Packet sizes | 64B, 256B, 512B, 1024B, 1500B |
| Avg. latency | 100, 150, 200us,1000us |

## 

## Packet mode



Number of packets per second (Mpps) per physical core (with siblings).

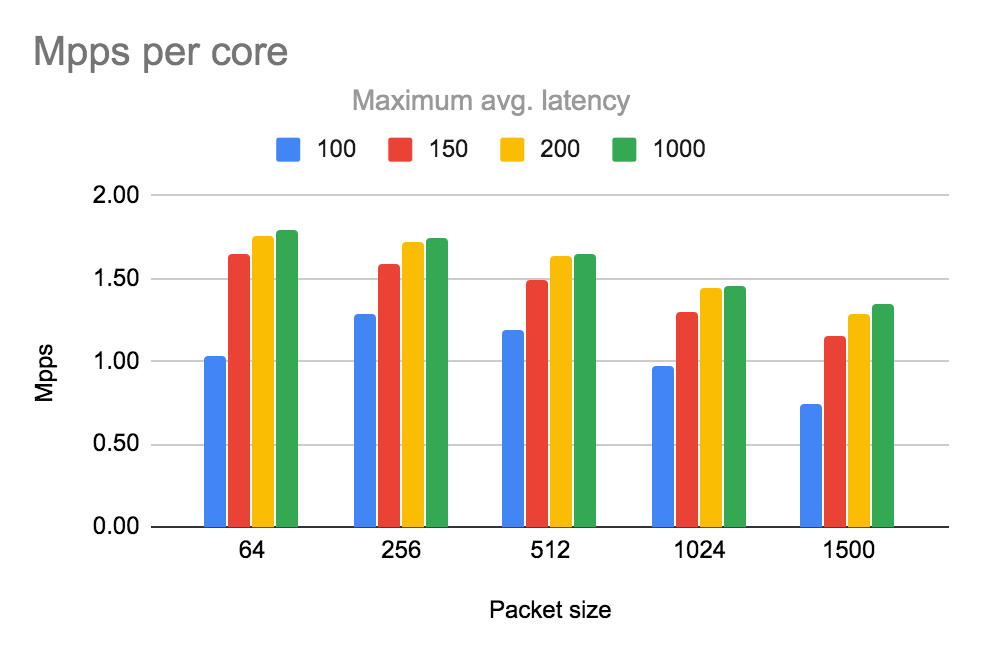
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Max Avg Latency [us] | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| 100 | 2.08 | 1.95 | 1.77 | 1.40 | 1.06 |
| 150 | 2.62 | 2.49 | 2.25 | 1.86 | 1.50 |
| 200 | 2.67 | 2.59 | 2.46 | 1.47 | 1.56 |
| 1000 | 2.76 | 2.66 | 2.49 | 2.19 | 1.58 |

# Chart

Number of gigabits per second (Gbps) per physical core (with siblings).

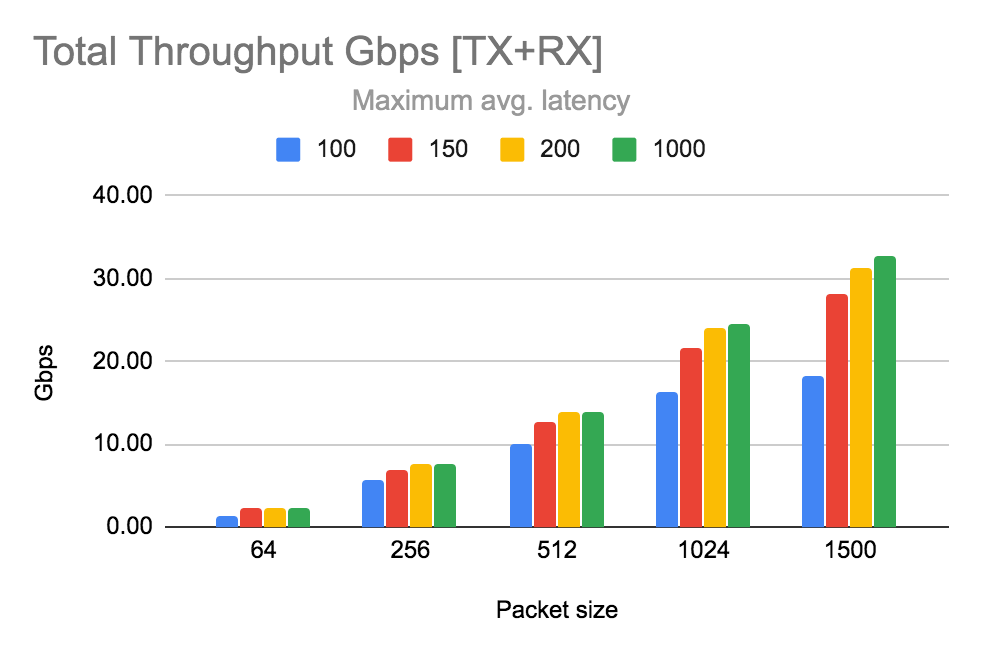
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Max Avg Latency [us] | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| 100 | 1.40 | 4.32 | 7.54 | 11.68 | 12.85 |
| 150 | 1.76 | 5.49 | 9.56 | 15.50 | 18.26 |
| 200 | 1.80 | 5.72 | 10.49 | 12.26 | 18.97 |
| 1000 | 1.90 | 5.90 | 10.60 | 18.30 | 19.20 |

## Flow mode



Number of packets per second (Mpps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Max Avg Latency [us] | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| 100 | 1.03 | 1.29 | 1.19 | 0.97 | 0.75 |
| 150 | 1.64 | 1.58 | 1.49 | 1.30 | 1.16 |
| 200 | 1.76 | 1.72 | 1.64 | 1.44 | 1.29 |
| 1000 | 1.79 | 1.74 | 1.65 | 1.46 | 1.34 |



Number of gigabits per second (Gbps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Max Avg Latency [us] | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| 100 | 1.39 | 5.68 | 10.10 | 16.23 | 18.18 |
| 150 | 2.21 | 6.99 | 12.66 | 21.68 | 28.12 |
| 200 | 2.36 | 7.60 | 13.95 | 24.06 | 31.29 |
| 1000 | 2.40 | 7.60 | 14.00 | 24.40 | 32.60 |

## 

## Summary

Reducing max average latency <200us impacts the performance (Mpps/core). Above 200us the limit is maximum packet drop 0.001%. Good practise is to not go with max Mpps/core and limit speed by 5% reducing latency 4 times as was presented in the case.

## 

## Intel x710 Fortville configuration

The NIC specific configuration parameters

|  |
| --- |
| cat /etc/sysconfig/network-scripts/ifcfg-vhost0  [...]  BIND\_INT=0000:19:00.0,0000:19:00.1  DRIVER=vfio-pci  BOND\_MODE=4  BOND\_POLICY=layer3+4  CPU\_LIST=2,4,30,32  SERVICE\_CORE\_MASK=0,1,28,29  DPDK\_CTRL\_THREAD\_MASK=0,1,28,29  sed -e "s/\x00/ /g" /proc/$(pidof contrail-vrouter-dpdk)/cmdline  /usr/bin/contrail-vrouter-dpdk --no-daemon --vr\_flow\_entries=2500000 --vr\_mempool\_sz 262144 --dpdk\_ctrl\_thread\_mask 0 --vr\_dpdk\_rx\_ring\_sz 2048 --vr\_dpdk\_tx\_ring\_sz 2048 --yield\_option 0 --service\_core\_mask (0,1,28,29) --dpdk\_ctrl\_thread\_mask (0,1,28,29) --socket-mem 1024 1024 --vdev eth\_bond\_bond1,mode=4,xmit\_policy=l34,socket\_id=0,mac=e4:43:4b:6e:70:b0,lacp\_rate=1,slave=0000:19:00.0,slave=0000:19:00.1 |

## 

## Detailed data

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **vRouter Mode** | **Data** | **Max avg latency** | **Packet size [bytes]** | | | | |
|  | **64** | **256** | **512** | **1024** | **1500** |
| Packet Mode | Total Throughput Gbps [TX+RX] | 100 | 2.79 | 8.63 | 15.08 | 23.36 | 25.70 |
| 150 | 3.52 | 10.98 | 19.12 | 31.00 | 36.52 |
| 200 | 3.59 | 11.45 | 20.98 | 24.51 | 37.95 |
| 1000 | 3.80 | 11.80 | 21.20 | 36.60 | 38.40 |
| Throughput Gbps/core | 100 | 1.40 | 4.32 | 7.54 | 11.68 | 12.85 |
| 150 | 1.76 | 5.49 | 9.56 | 15.50 | 18.26 |
| 200 | 1.80 | 5.72 | 10.49 | 12.26 | 18.97 |
| 1000 | 1.90 | 5.90 | 10.60 | 18.30 | 19.20 |
| Total Mpps [TX+RX] | 100 | 4.16 | 3.91 | 3.54 | 2.80 | 2.11 |
| 150 | 5.23 | 4.97 | 4.49 | 3.71 | 3.00 |
| 200 | 5.35 | 5.18 | 4.93 | 2.93 | 3.12 |
| 1000 | 5.52 | 5.33 | 4.99 | 4.39 | 3.16 |
| Mpps/core | 100 | 2.08 | 1.95 | 1.77 | 1.40 | 1.06 |
| 150 | 2.62 | 2.49 | 2.25 | 1.86 | 1.50 |
| 200 | 2.67 | 2.59 | 2.46 | 1.47 | 1.56 |
| 1000 | 2.76 | 2.66 | 2.49 | 2.19 | 1.58 |
| Avg. Latency | 100 | 99.00 | 99 | 99 | 99 | 99 |
| 150 | 149 | 149 | 149 | 149 | 149 |
| 200 | 178 | 194 | 198 | 103 | 199 |
| 1000 | 831 | 578 | 461 | 184 | 134 |
| Flow Mode | Total Throughput Gbps [TX+RX] | 100 | 1.39 | 5.68 | 10.10 | 16.23 | 18.18 |
| 150 | 2.21 | 6.99 | 12.66 | 21.68 | 28.12 |
| 200 | 2.36 | 7.60 | 13.95 | 24.06 | 31.29 |
| 1000 | 2.40 | 7.60 | 14.00 | 24.40 | 32.60 |
| Throughput Gbps/core | 100 | 0.69 | 2.84 | 5.05 | 8.12 | 9.09 |
| 150 | 1.10 | 3.50 | 6.33 | 10.84 | 14.06 |
| 200 | 1.18 | 3.80 | 6.97 | 12.03 | 15.64 |
| 1000 | 1.20 | 3.80 | 7.00 | 12.20 | 16.30 |
| Total Mpps [TX+RX] | 100 | 2.06 | 2.57 | 2.37 | 1.94 | 1.50 |
| 150 | 3.28 | 3.17 | 2.97 | 2.60 | 2.31 |
| 200 | 3.52 | 3.44 | 3.28 | 2.88 | 2.57 |
| 1000 | 3.57 | 3.49 | 3.30 | 2.92 | 2.68 |
| Mpps/core | 100 | 1.03 | 1.29 | 1.19 | 0.97 | 0.75 |
| 150 | 1.64 | 1.58 | 1.49 | 1.30 | 1.16 |
| 200 | 1.76 | 1.72 | 1.64 | 1.44 | 1.29 |
| 1000 | 1.79 | 1.74 | 1.65 | 1.46 | 1.34 |
| Avg. Latency | 100 | 77 | 99 | 99 | 99 | 99 |
| 150 | 149 | 149 | 149 | 149 | 149 |
| 200 | 192 | 199 | 199 | 199 | 199 |
| 1000 | 323 | 292 | 256 | 218 | 242 |

## 

# Test case 3: Number of packets per second (Mpps) and Throughput (Gbps) per core on Intel XXV710 Ethernet Adapters

## Configuration of vRouter on compute-B (DUT):

|  |  |
| --- | --- |
| Operating System | RedHat 7.7 |
| Number of Forwarding Cores | 2 (Physical) + 2 (Siblings) |
| Drop Rate | 0.001% |
| Bond | 2 x 10G\* (802.3ad LACP) with L3/L4 hash |
| Hugepages | 1G |
| NUMA allocation | vRouter, NIC and VM on the same socket (NUMA 0) |

\*NICs are 25G capable but due the bug CEM-13470 don’t work

## CPU allocation (Full-CPU partitioning scheme) on compute-B:

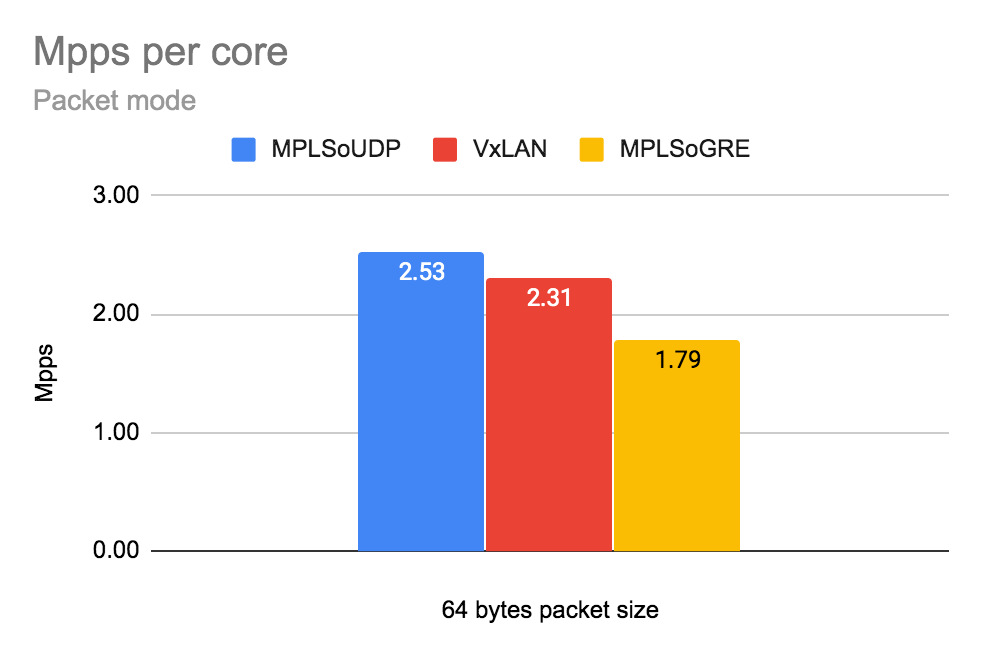
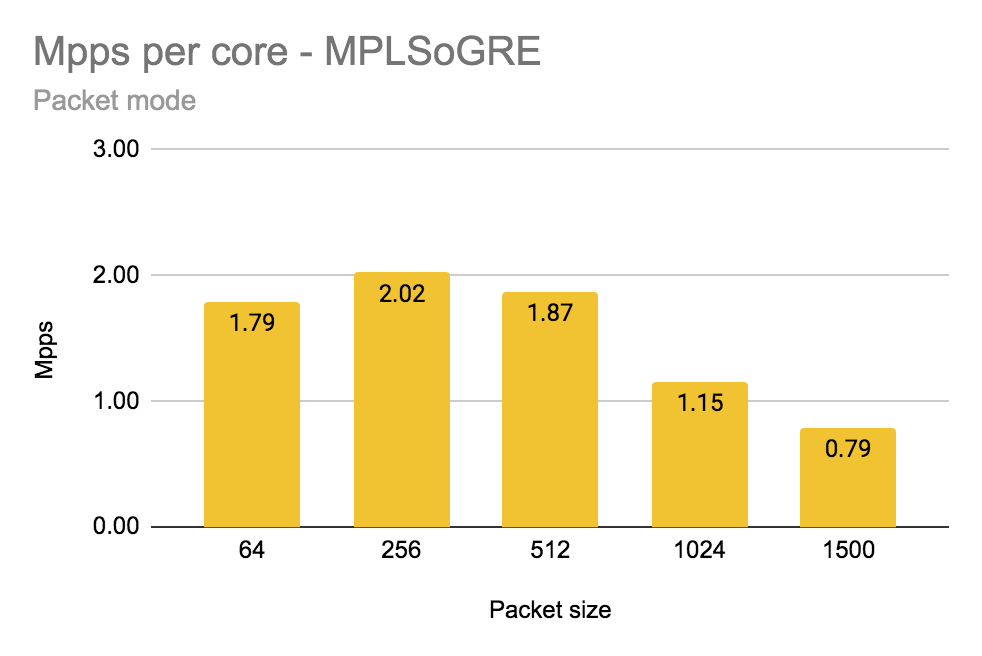
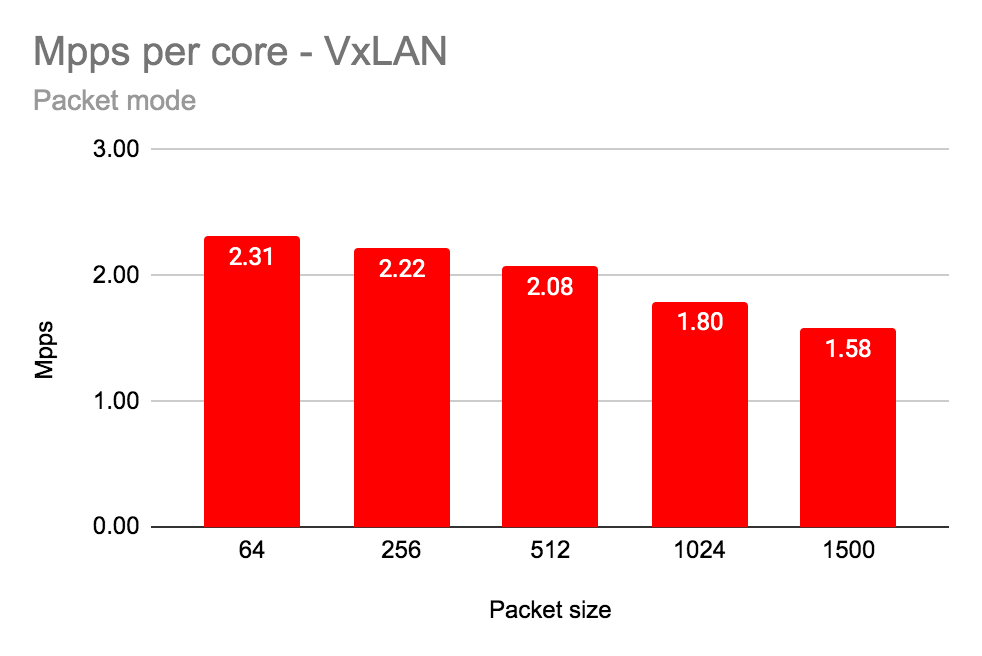
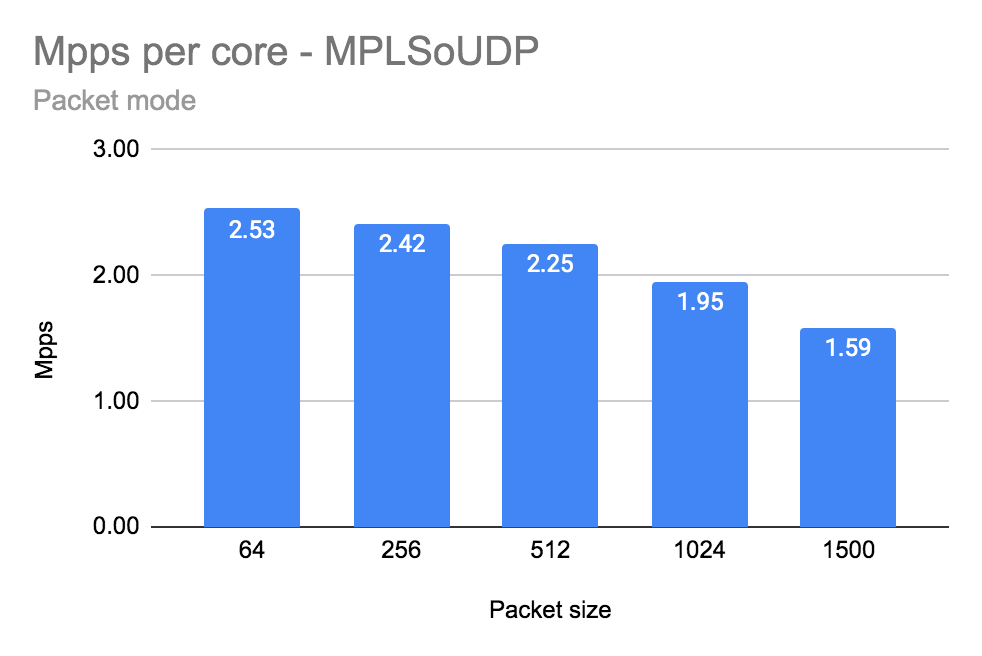
|  |  |
| --- | --- |
| vRouter forwarding cores | 2,4,30,32 |
| Service cores | 0,1,28,29 |
| Control cores | 0,1,28,29 |
| Nova (VM) cores [VM core: HV core] | 0: 18,1: 46, 2: 40, 3: 12, 4: 48, 5: 20, 6: 24, 7: 52, 8: 10, 9: 38 |
| Host OS cores | 0,1,28,29 |
| Kernel isolcpus and tuned isolcpu list | 2-27,30-55 |

## 

## Traffic pattern

|  |  |
| --- | --- |
| Number of flows | 16535 |
| Encapsulation | MPLSoUDP, MPLSoGRE, VxLAN |
| Packet size | 64B, 256B, 512B, 1024B, 1500B |

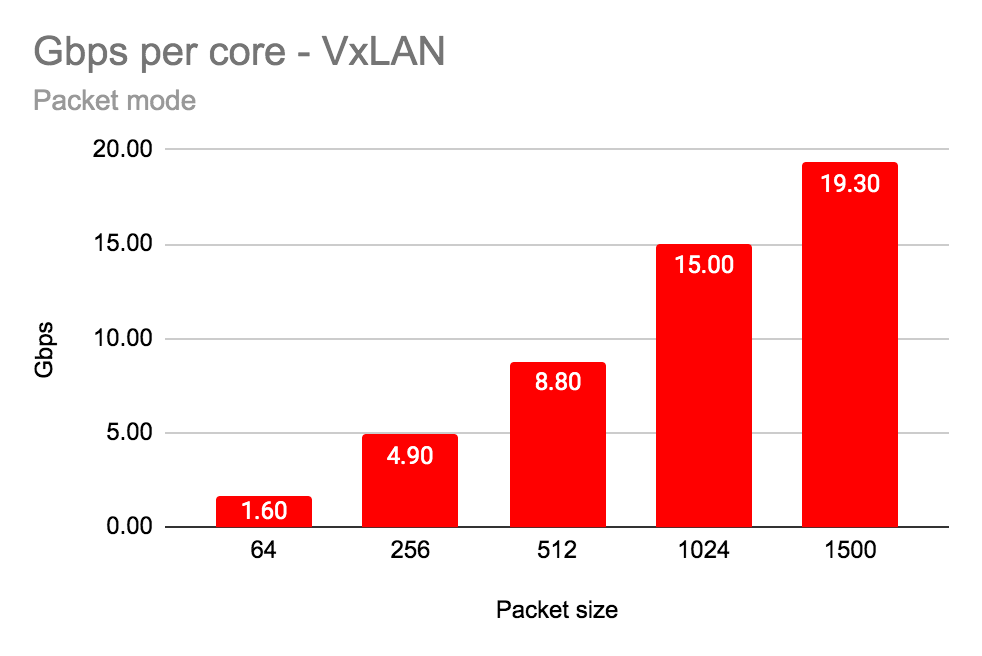
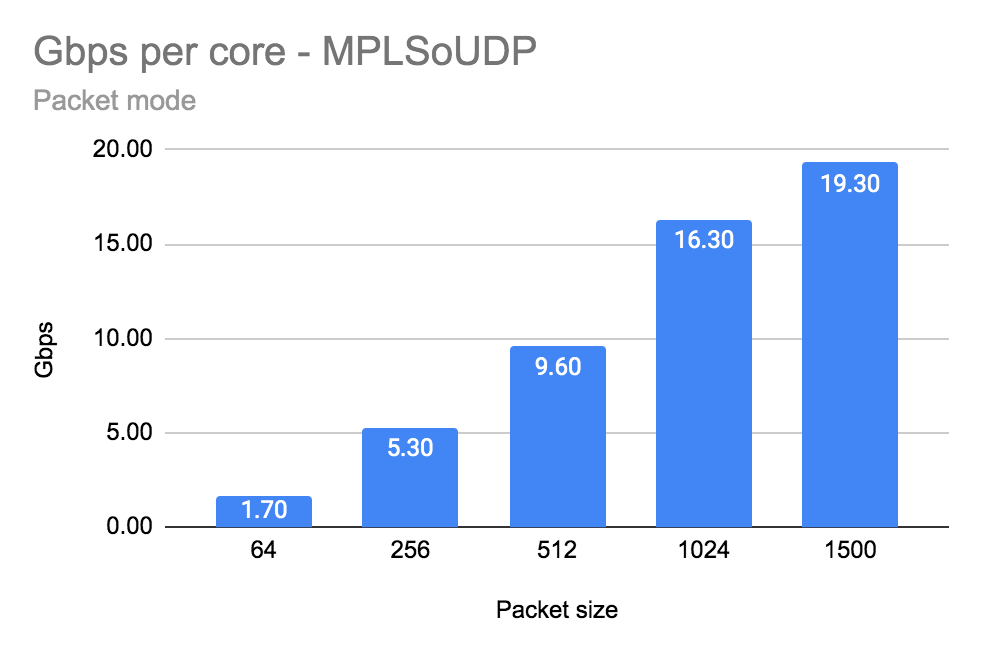
## Packet mode (Mpps):

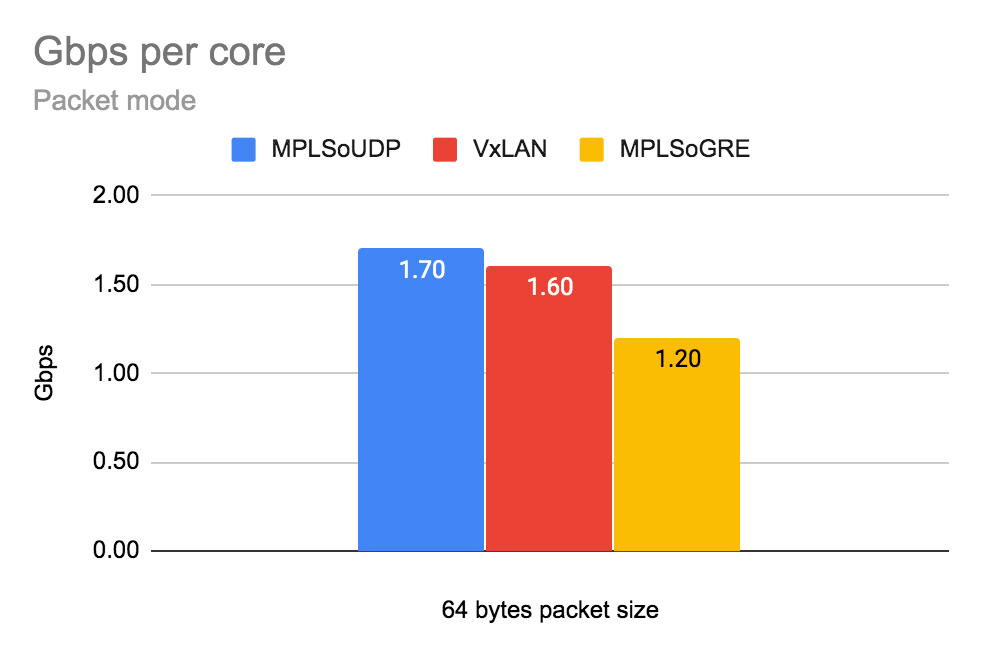
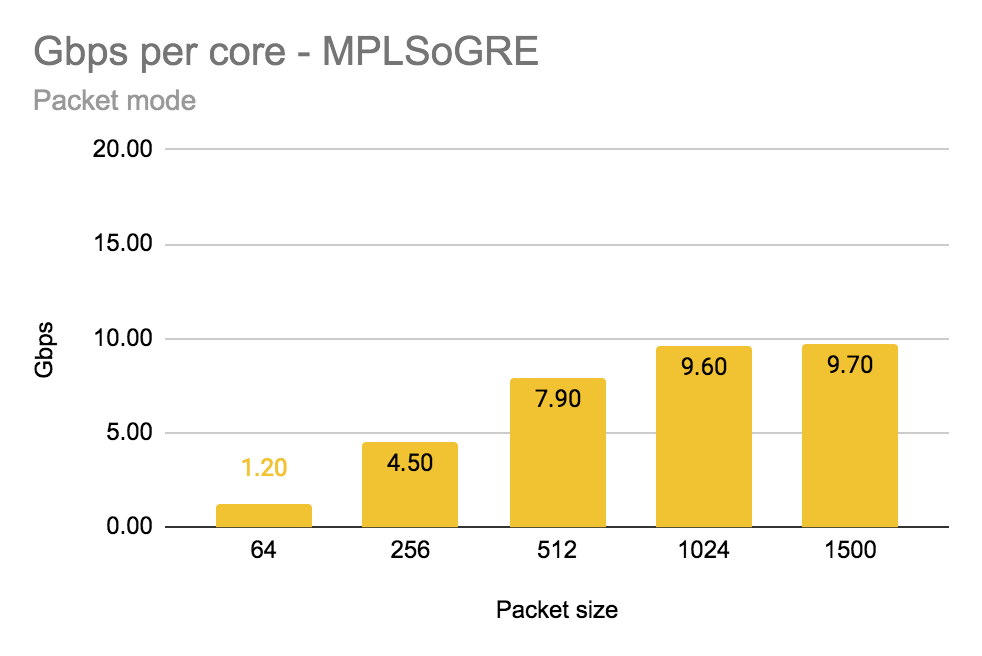


Number of packets per second (Mpps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Encapsulation | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| MPLSoUDP | 2.53 | 2.42 | 2.25 | 1.95 | 1.59 |
| MPLSoGRE | 1.79 | 2.02 | 1.87 | 1.15 | 0.79 |
| VxLAN | 2.31 | 2.22 | 2.08 | 1.80 | 1.58 |

## Packet mode (Gbps):

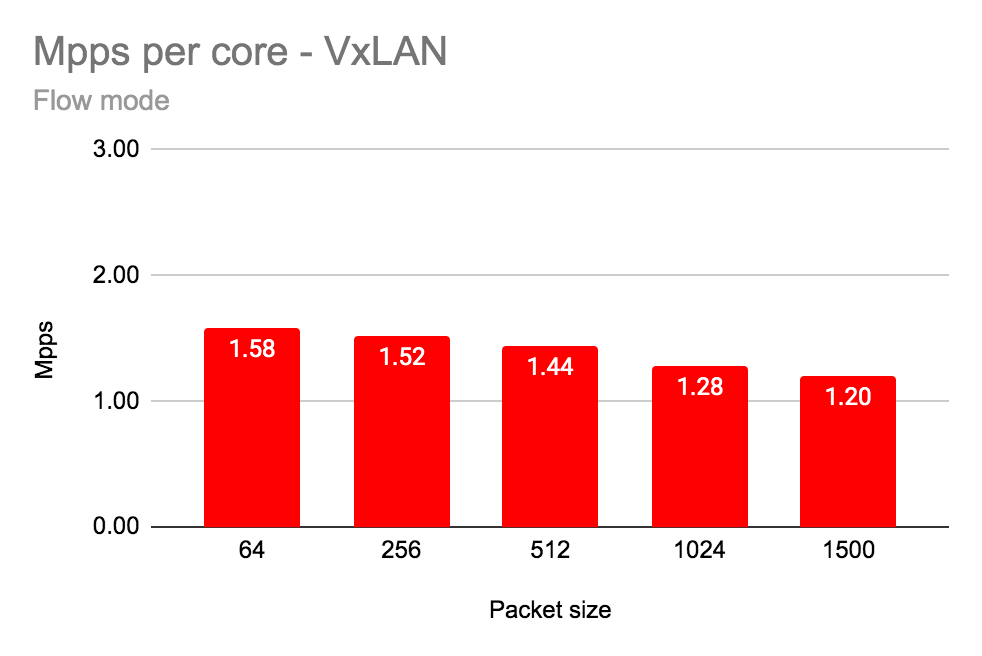
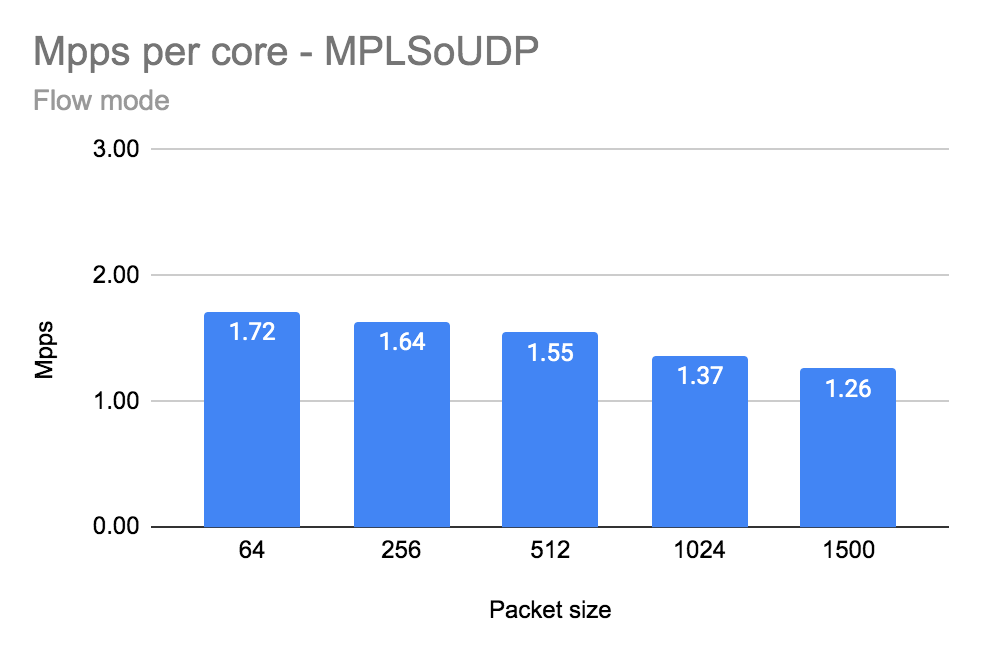


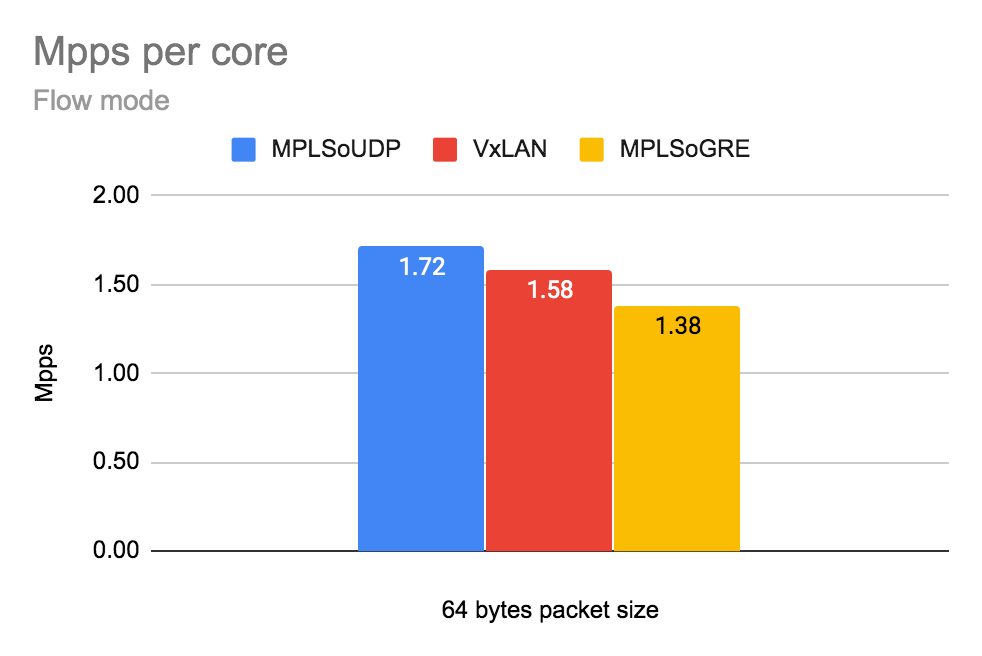
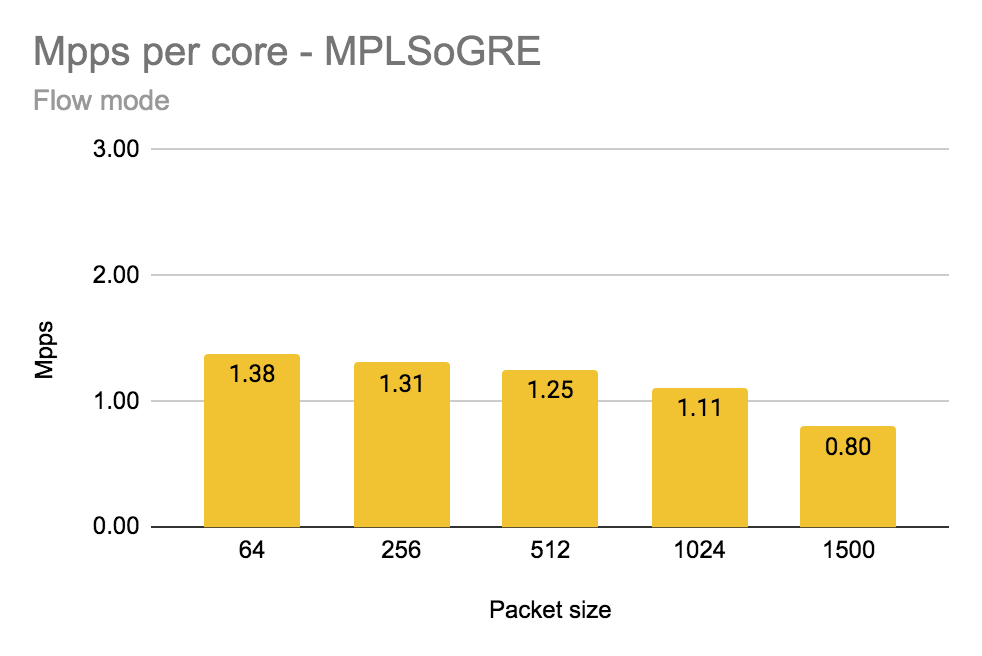


Number of gigabits per second (Gbps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Encapsulation | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| MPLSoUDP | 1.70 | 5.30 | 9.60 | 16.30 | 19.30 |
| MPLSoGRE | 1.20 | 4.50 | 7.90 | 9.60 | 9.70 |
| VxLAN | 1.60 | 4.90 | 8.80 | 15.00 | 19.30 |

## Flow mode (Mpps):

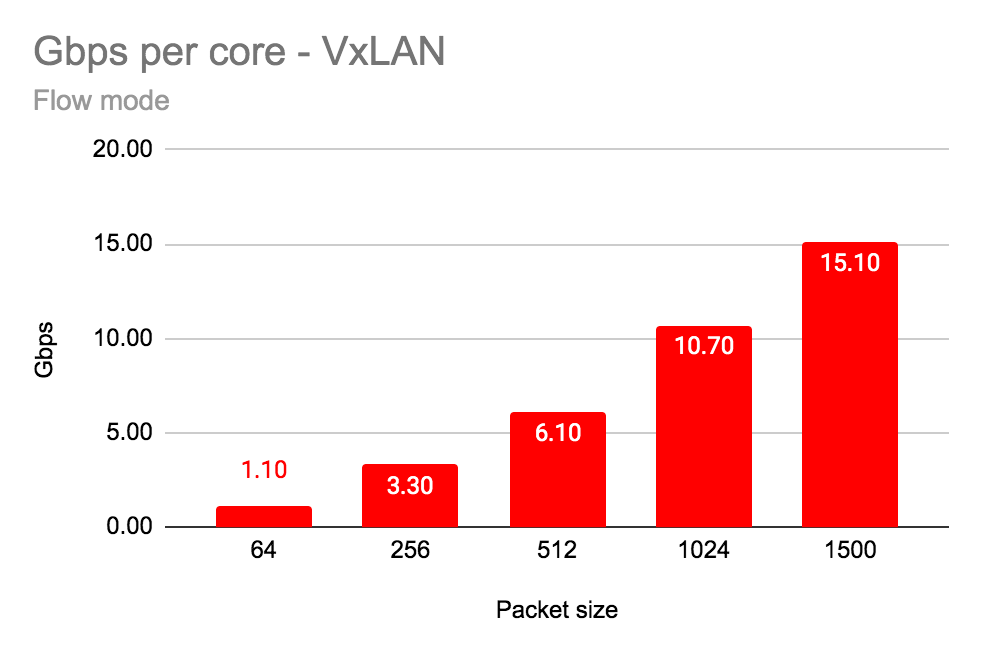
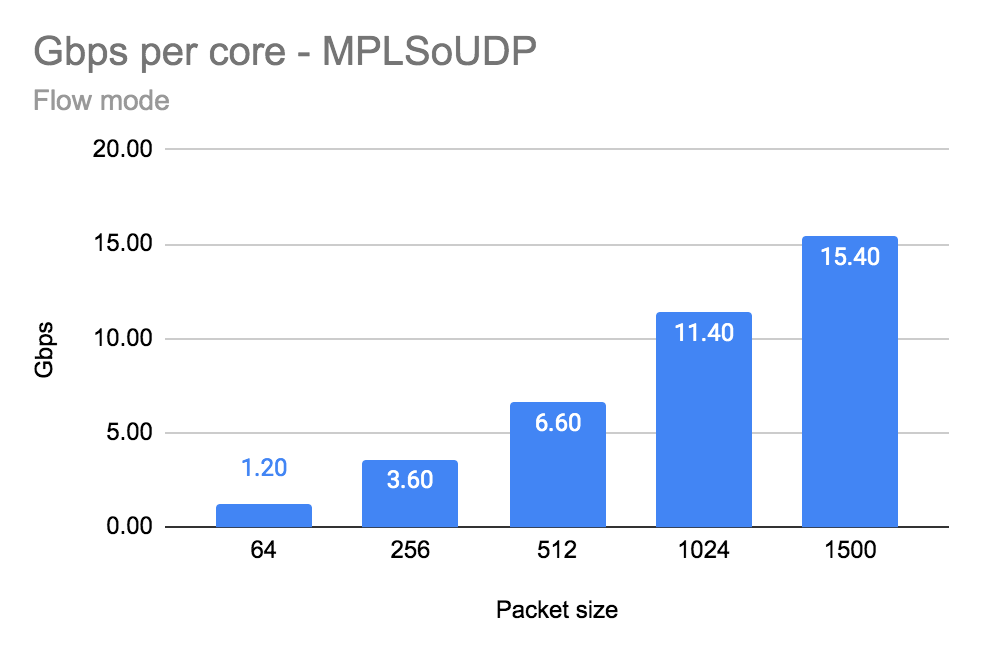


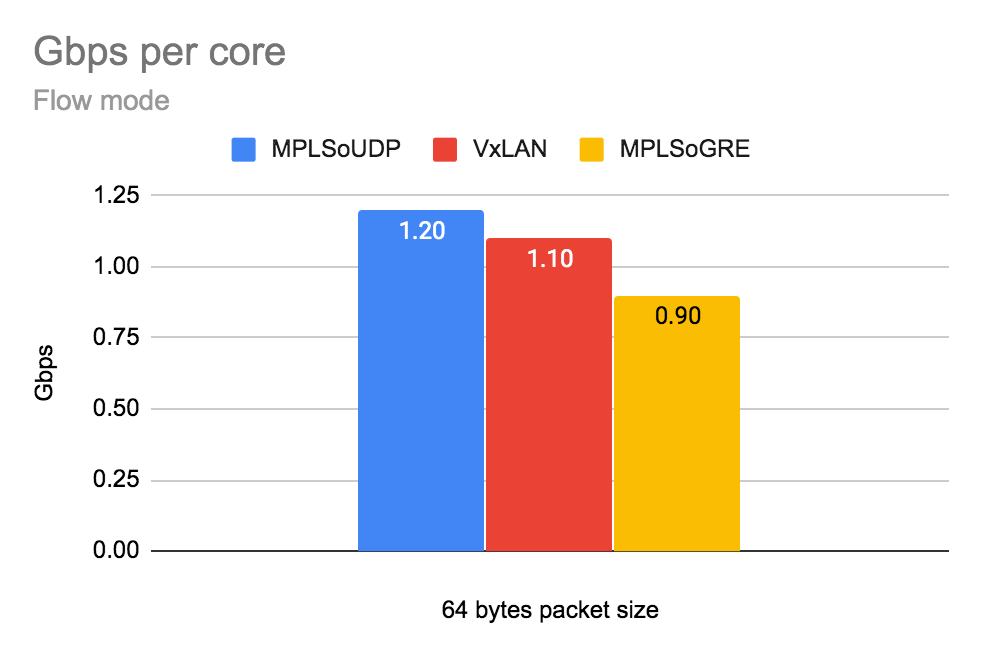
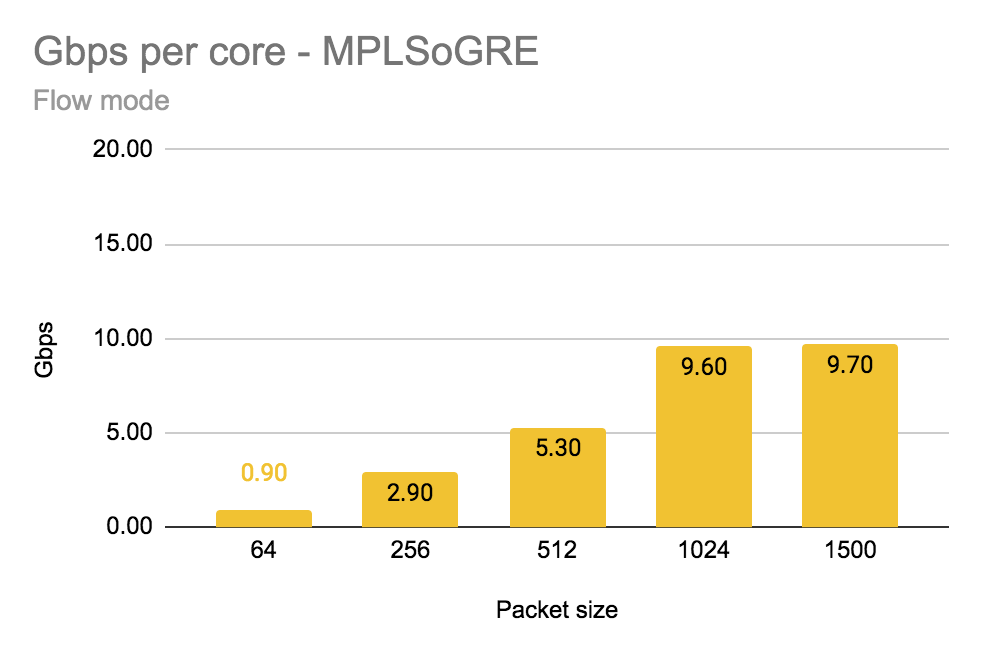


Number of packets per second (Mpps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Encapsulation | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| MPLSoUDP | 1.72 | 1.64 | 1.55 | 1.37 | 1.26 |
| MPLSoGRE | 1.38 | 1.31 | 1.25 | 1.11 | 0.80 |
| VxLAN | 1.58 | 1.52 | 1.44 | 1.28 | 1.20 |

## Flow mode (Gbps):





Number of gigabits per second (Gbps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Encapsulation | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| MPLSoUDP | 1.00 | 3.60 | 6.60 | 11.40 | 15.40 |
| MPLSoGRE | 0.90 | 2.90 | 5.30 | 9.60 | 9.70 |
| VxLAN | 1.10 | 3.30 | 6.10 | 10.70 | 15.10 |

## 

## 

## Summary

On a production with bursty traffic it is recommended to configure larger buffer descriptors size (RX/TX) towards NIC cards using parameters: *dpdk\_txd\_sz and dpdk\_rxd\_sz and adjust memory pool size* vr\_mempool\_sz (default value is 128).

The formula to calculate vr\_mempool\_sz:

**mempool size** = *2 \* (number\_of\_RX\_descriptors + number\_of\_TX\_descriptors) \* number\_of\_vrouter\_cores \* number\_of\_ports\_in\_dpdk\_bond*

## Intel XXV710 Fortville configuration

The NIC specific configuration parameters

|  |
| --- |
| cat /etc/sysconfig/network-scripts/ifcfg-vhost0  [...]  #xxv710  BIND\_INT=0000:d8:00.0,0000:d8:00.1  DRIVER=vfio-pci  BOND\_MODE=4  BOND\_POLICY=layer3+4  CPU\_LIST=2,4,30,32  ps aux | grep "/usr/bin/contrail-vrouter-dpdk --no-daemon"  root 522803 407 0.0 445316052 11472 ? SLl 11:23 231:48 /usr/bin/contrail-vrouter-dpdk --no-daemon --vr\_flow\_entries=2500000 --vr\_mempool\_sz 262144 --vr\_dpdk\_rx\_ring\_sz 2048 --vr\_dpdk\_tx\_ring\_sz 2048 --yield\_option 0 --service\_core\_mask (0,1,28,29) --dpdk\_ctrl\_thread\_mask (0,1,28,29) --socket-mem 1024 1024 --vdev eth\_bond\_bond1,mode=4,xmit\_policy=l34,socket\_id=0,mac=3c:fd:fe:eb:1d:2c,lacp\_rate=1,slave=0000:d8:00.0,slave=0000:d8:00.1 |

## 

## 

## Detailed data

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **vRouter Mode** | **Data** | **Encapsulation** | **Frame size [bytes]** | | | | |
|  | **64** | **256** | **512** | **1024** | **1500** |
| Packet Mode | Total Throughput Gbps [TX+RX] | MPLSoUDP | 3.40 | 10.60 | 19.20 | 32.60 | 38.60 |
| MPLSoGRE | 2.40 | 9.00 | 15.80 | 19.20 | 19.40 |
| VxLAN | 3.20 | 9.80 | 17.60 | 30.00 | 38.60 |
| Throughput Gbps/core | MPLSoUDP | 1.70 | 5.30 | 9.60 | 16.30 | 19.30 |
| MPLSoGRE | 1.20 | 4.50 | 7.90 | 9.60 | 9.70 |
| VxLAN | 1.60 | 4.90 | 8.80 | 15.00 | 19.30 |
| Total Mpps [TX+RX] | MPLSoUDP | 5.06 | 4.83 | 4.51 | 3.91 | 3.18 |
| MPLSoGRE | 3.57 | 4.04 | 3.73 | 2.31 | 1.58 |
| VxLAN | 4.62 | 4.43 | 4.15 | 3.59 | 3.17 |
| Mpps/core | MPLSoUDP | 2.53 | 2.42 | 2.25 | 1.95 | 1.59 |
| MPLSoGRE | 1.79 | 2.02 | 1.87 | 1.15 | 0.79 |
| VxLAN | 2.31 | 2.22 | 2.08 | 1.80 | 1.58 |
| Avg. Latency | MPLSoUDP | 871 | 775 | 565 | 297 | 265 |
| MPLSoGRE | 98 | 435 | 499 | 121 | 99 |
| VxLAN | 484 | 603 | 415 | 261 | 316 |
| Flow Mode | Total Throughput Gbps [TX+RX] | MPLSoUDP | 2.40 | 7.20 | 13.20 | 22.80 | 30.80 |
| MPLSoGRE | 1.80 | 5.80 | 10.60 | 19.20 | 19.40 |
| VxLAN | 2.20 | 6.60 | 12.20 | 21.40 | 30.20 |
| Throughput Gbps/core | MPLSoUDP | 1.20 | 3.60 | 6.60 | 11.40 | 15.40 |
| MPLSoGRE | 0.90 | 2.90 | 5.30 | 9.60 | 9.70 |
| VxLAN | 1.10 | 3.30 | 6.10 | 10.70 | 15.10 |
| Total Mpps [TX+RX] | MPLSoUDP | 3.43 | 3.27 | 3.09 | 2.73 | 2.53 |
| MPLSoGRE | 2.76 | 2.61 | 2.51 | 2.23 | 1.60 |
| VxLAN | 3.17 | 3.03 | 2.88 | 2.55 | 2.40 |
| Mpps/core | MPLSoUDP | 1.72 | 1.64 | 1.55 | 1.37 | 1.26 |
| MPLSoGRE | 1.38 | 1.31 | 1.25 | 1.11 | 0.80 |
| VxLAN | 1.58 | 1.52 | 1.44 | 1.28 | 1.20 |
| Avg. Latency | MPLSoUDP | 458 | 360 | 344 | 276 | 276 |
| MPLSoGRE | 129 | 687 | 633 | 191 | 142 |
| VxLAN | 208 | 236 | 237 | 243 | 281 |

# Test case 4: Compare NICs results Intel 82599 vs Intel X710 vs Intel XXV710

# 

## Configuration of vRouter on compute-B (DUT):

|  |  |
| --- | --- |
| Operating System | RedHat 7.7 |
| Number of Forwarding Cores | 2 (Physical) + 2 (Siblings) |
| Drop Rate | 0.001% |
| Bond | 2 x 10G (802.3ad LACP) with L3/L4 hash |
| Hugepages | 1G |
| NUMA allocation | vRouter, NIC and VM on the same socket (NUMA 0) |

## 

## CPU allocation (Full-CPU partitioning scheme) on compute-B:

|  |  |
| --- | --- |
| vRouter forwarding cores | 2,4,30,32 |
| Service cores | 0,1,28,29 |
| Control cores | 0,1,28,29 |
| Nova (VM) cores [VM core: HV core] | 0: 18,1: 46, 2: 40, 3: 12, 4: 48, 5: 20, 6: 24, 7: 52, 8: 10, 9: 38 |
| Host OS cores | 0,1,28,29 |
| Kernel isolcpus and tuned isolcpu list | 2-27,30-55 |

## 

## Traffic pattern

|  |  |
| --- | --- |
| Number of flows | 16535 |
| Encapsulation | MPLSoUDP, MPLSoGRE, VxLAN |
| Packet size | 64B, 256B, 512B, 1024B, 1500B |

## 

## Packet mode:

## ChartChart

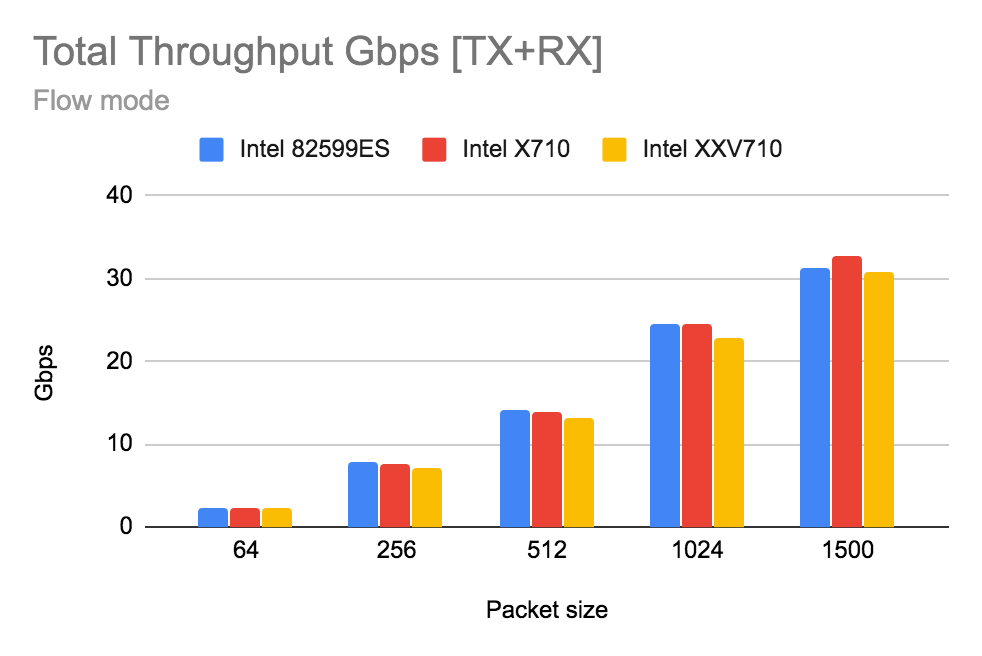
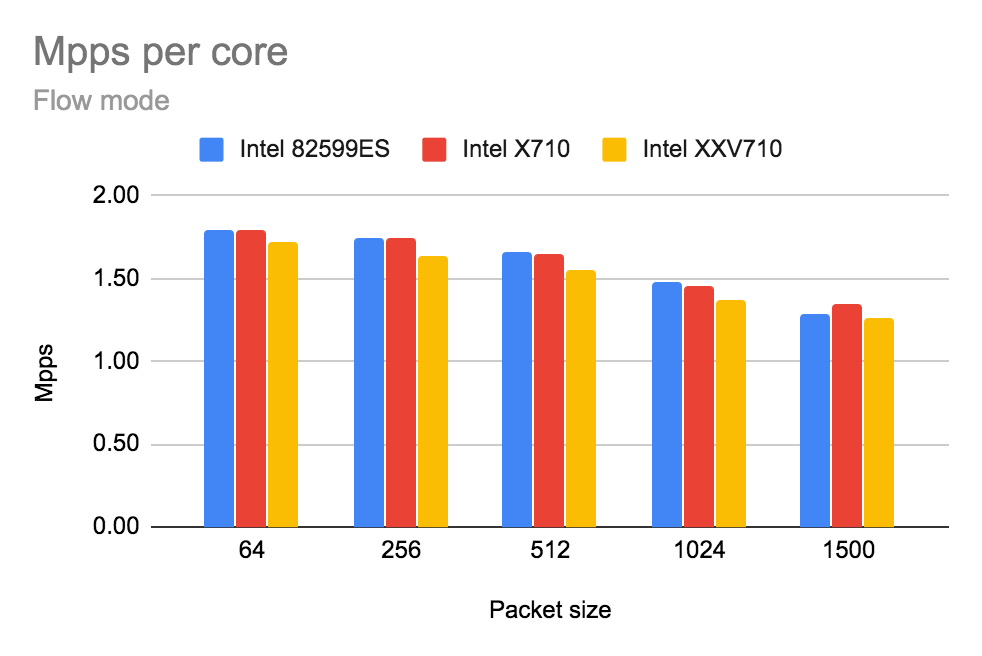
Number of packets per second (Mpps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Network card | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| Intel 82599ES | 2.75 | 2.66 | 2.33 | 1.66 | 1.31 |
| Intel X710 | 2.76 | 2.66 | 2.49 | 2.19 | 1.58 |
| Intel XXV710 | 2.53 | 2.42 | 2.25 | 1.95 | 1.59 |

Total number of gigabits [TX+RX] per second (Gbps).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Network card | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| Intel 82599ES | 3.60 | 11.80 | 19.80 | 27.80 | 31.80 |
| Intel X710 | 3.80 | 11.80 | 21.20 | 36.60 | 38.40 |
| Intel XXV710 | 3.40 | 10.60 | 19.20 | 32.60 | 38.60 |

## Flow mode:



Number of packets per second (Mpps) per physical core (with siblings).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Network card | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| Intel 82599ES | 1.79 | 1.75 | 1.66 | 1.47 | 1.28 |
| Intel X710 | 1.79 | 1.74 | 1.65 | 1.46 | 1.34 |
| Intel XXV710 | 1.72 | 1.64 | 1.55 | 1.37 | 1.26 |

Total number of gigabits [TX+RX] per second (Gbps).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Network card | Packet Size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| Intel 82599ES | 2.40 | 7.80 | 14.20 | 24.60 | 31.20 |
| Intel X710 | 2.40 | 7.60 | 14.00 | 24.40 | 32.60 |
| Intel XXV710 | 2.40 | 7.20 | 13.20 | 22.80 | 30.80 |

## 

## Detailed data

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **vRouter Mode** | **Data** | **Encapsulation** | **Packet size [bytes]** | | | | |
|  | **64** | **256** | **512** | **1024** | **1500** |
| Packet Mode | Total Throughput Gbps [TX+RX] | Intel 82599ES | 3.60 | 11.80 | 19.80 | 27.80 | 31.80 |
| Intel X710 | 3.80 | 11.80 | 21.20 | 36.60 | 38.40 |
| Intel XXV710 | 3.40 | 10.60 | 19.20 | 32.60 | 38.60 |
| Throughput Gbps/core | Intel 82599ES | 1.80 | 5.90 | 9.90 | 13.90 | 15.90 |
| Intel X710 | 1.90 | 5.90 | 10.60 | 18.30 | 19.20 |
| Intel XXV710 | 1.70 | 5.30 | 9.60 | 16.30 | 19.30 |
| Total Mpps [TX+RX] | Intel 82599ES | 5.49 | 5.32 | 4.65 | 3.33 | 2.62 |
| Intel X710 | 5.52 | 5.33 | 4.99 | 4.39 | 3.16 |
| Intel XXV710 | 5.06 | 4.83 | 4.51 | 3.91 | 3.18 |
| Mpps/core | Intel 82599ES | 2.75 | 2.66 | 2.33 | 1.66 | 1.31 |
| Intel X710 | 2.76 | 2.66 | 2.49 | 2.19 | 1.58 |
| Intel XXV710 | 2.53 | 2.42 | 2.25 | 1.95 | 1.59 |
| Avg. Latency | Intel 82599ES | 862 | 628 | 191 | 142 | 130 |
| Intel X710 | 718 | 487 | 379 | 286 | 228 |
| Intel XXV710 | 871 | 775 | 565 | 297 | 265 |
| Flow Mode | Total Throughput Gbps [TX+RX] | Intel 82599ES | 2.40 | 7.80 | 14.20 | 24.60 | 31.20 |
| Intel X710 | 2.40 | 7.60 | 14.00 | 24.40 | 32.60 |
| Intel XXV710 | 2.40 | 7.20 | 13.20 | 22.80 | 30.80 |
| Throughput Gbps/core | Intel 82599ES | 1.20 | 3.90 | 7.10 | 12.30 | 15.60 |
| Intel X710 | 1.20 | 3.80 | 7.00 | 12.20 | 16.30 |
| Intel XXV710 | 1.20 | 3.60 | 6.60 | 11.40 | 15.40 |
| Total Mpps [TX+RX] | Intel 82599ES | 3.57 | 3.49 | 3.31 | 2.95 | 2.56 |
| Intel X710 | 3.57 | 3.49 | 3.30 | 2.92 | 2.68 |
| Intel XXV710 | 3.43 | 3.27 | 3.09 | 2.73 | 2.53 |
| Mpps/core | Intel 82599ES | 1.79 | 1.75 | 1.66 | 1.47 | 1.28 |
| Intel X710 | 1.79 | 1.74 | 1.65 | 1.46 | 1.34 |
| Intel XXV710 | 1.72 | 1.64 | 1.55 | 1.37 | 1.26 |
| Avg. Latency | Intel 82599ES | 271 | 321 | 290 | 257 | 233 |
| Intel X710 | 323 | 292 | 256 | 218 | 242 |
| Intel XXV710 | 458 | 360 | 344 | 276 | 276 |

# Test case 5: Number of packets per second (Mpps) and Throughput (Gbps) per core with increasing number of vRouter cores

## Configuration of vRouter on compute-B (DUT):

|  |  |
| --- | --- |
| Operating System | RedHat 7.7 |
| Number of Forwarding Cores | various |
| Drop Rate | 0.001% |
| NIC | Intel x710 |
| Bond | 2 x 10G (802.3ad LACP) with L3/L4 hash |
| Hugepages | 1G |
| NUMA allocation | vRouter, NIC and VM on the same socket (NUMA 0) |

## CPU allocation (Full-CPU partitioning scheme) on compute-B:

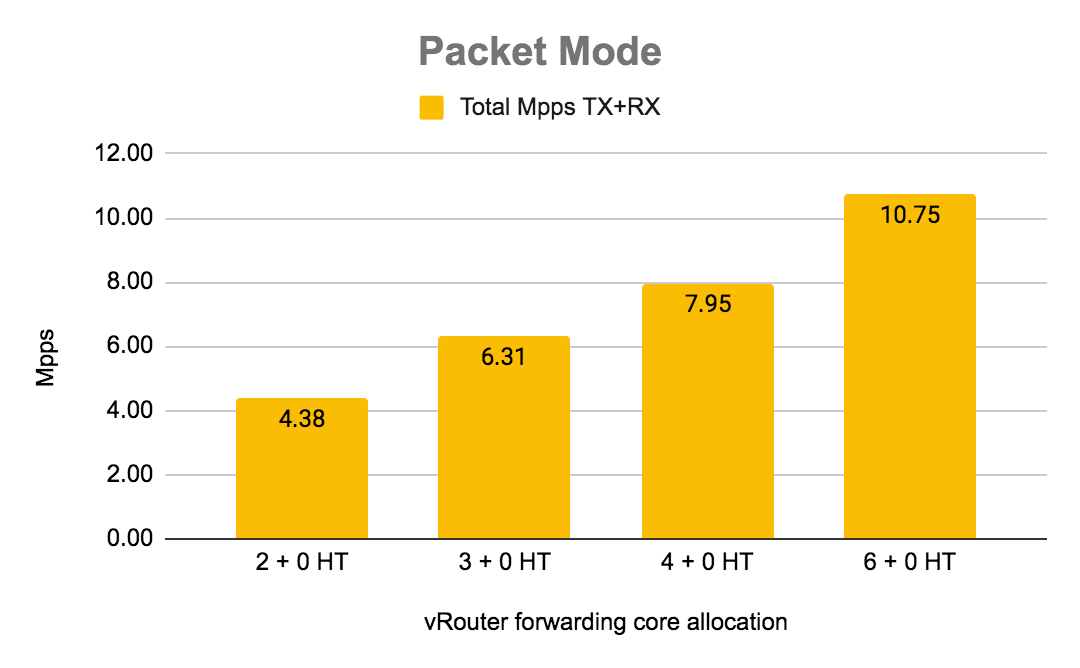
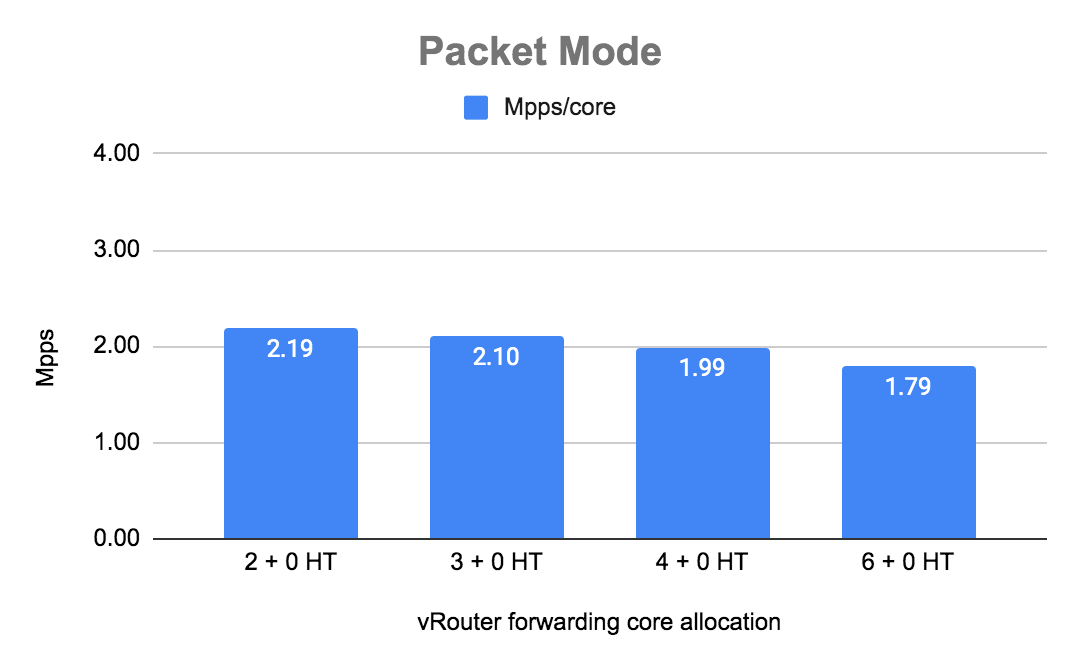
|  |  |
| --- | --- |
| vRouter forwarding cores [scenario:allocated cores] | 1+1HT: 2,30; 2+0HT: 2,4; 2+2HT: 2,4,30,32; 3+0HT: 2,4,6; 3+3HT: 2,4,6,30,32,34; 4+0HT: 2,4,6,8; 4+4HT:2,4,6,8,30,32,34,36 |
| Service cores | 0,1,28,29 |
| Control cores | 0,1,28,29 |
| Nova (VM) cores [VM core: HV core] | 0: 18,1: 46, 2: 40, 3: 12, 4: 48, 5: 20, 6: 24, 7: 52, 8: 10, 9: 38 |
| Host OS cores | 0,1,28,29 |
| Kernel isolcpus and tuned isolcpu list | 2-27,30-55 |

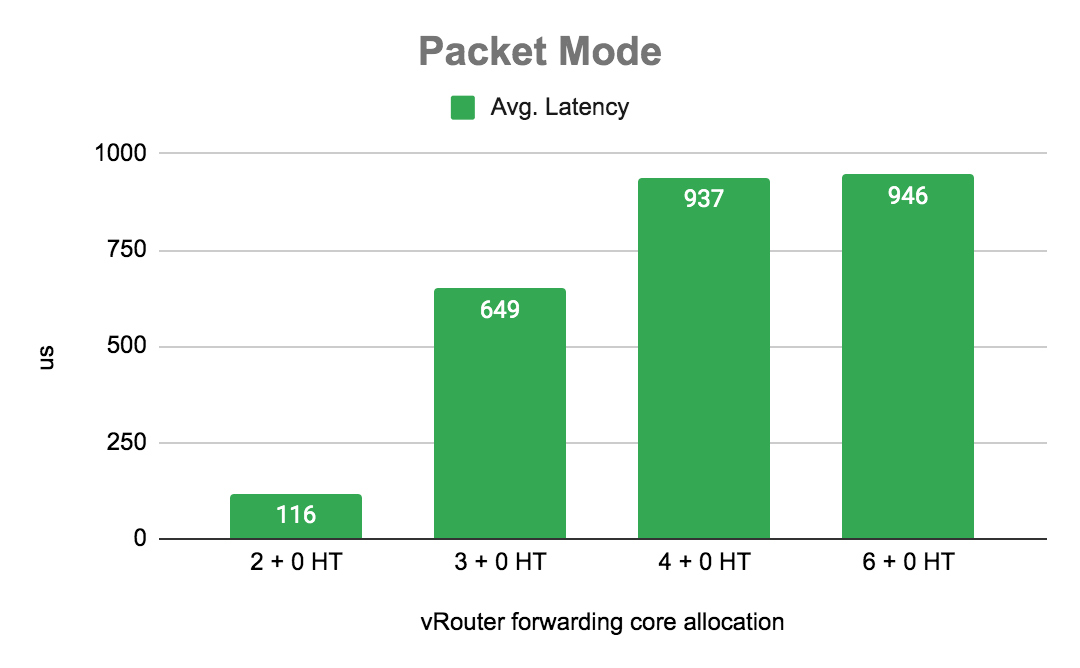
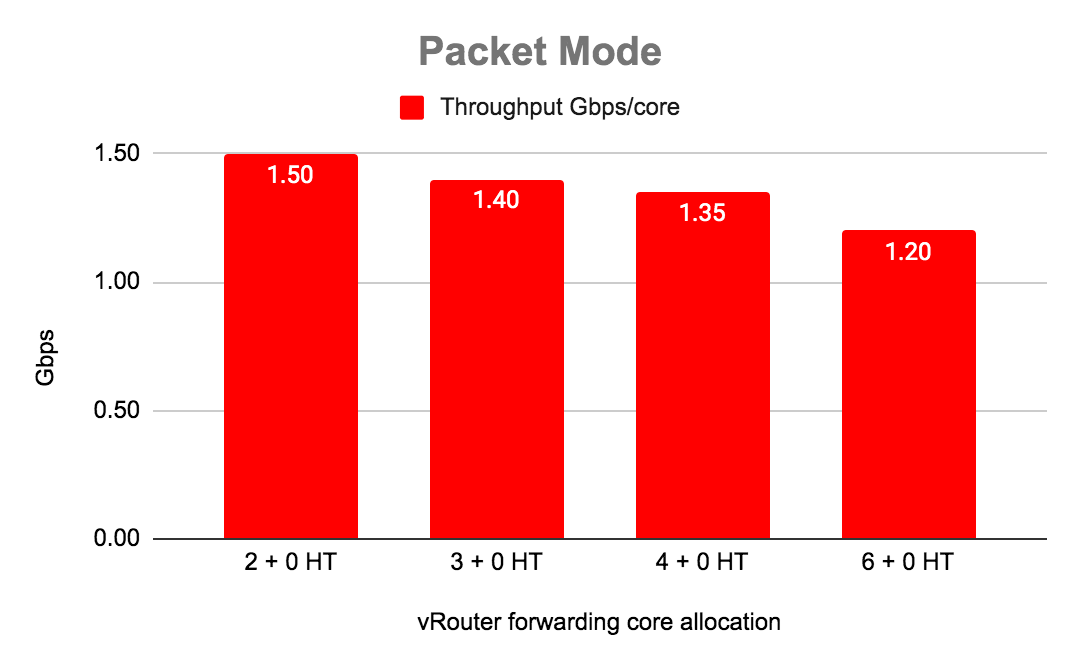
## Traffic pattern

|  |  |
| --- | --- |
| Number of flows | 16535 |
| Encapsulation | MPLSoUDP |
| Packet size | 64B |

## 

## Packet mode without siblings (HT):

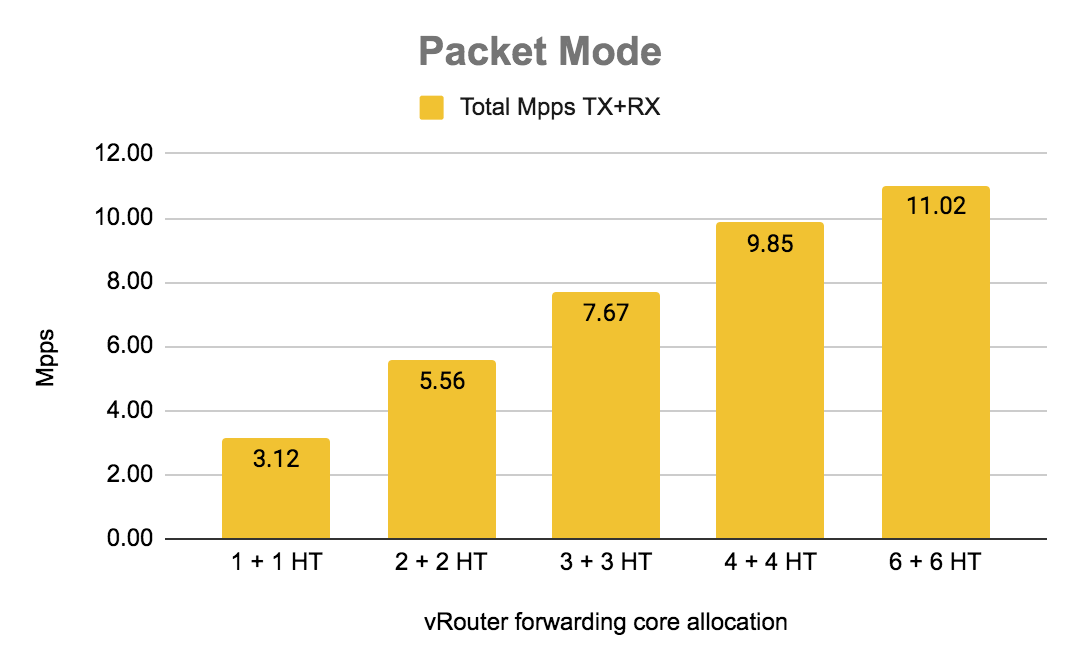
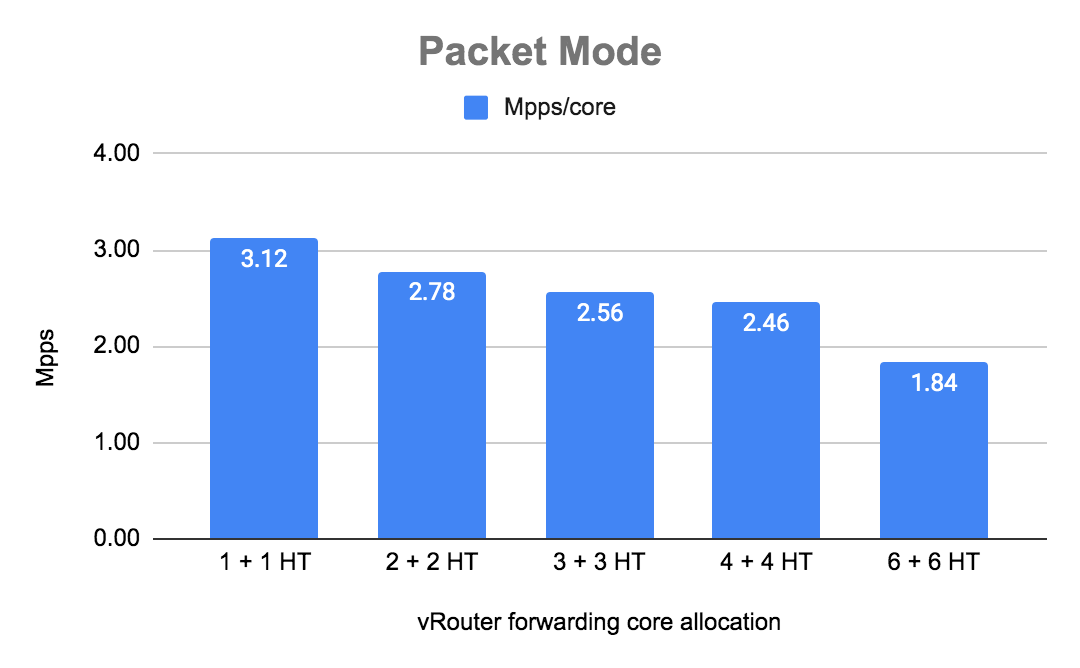


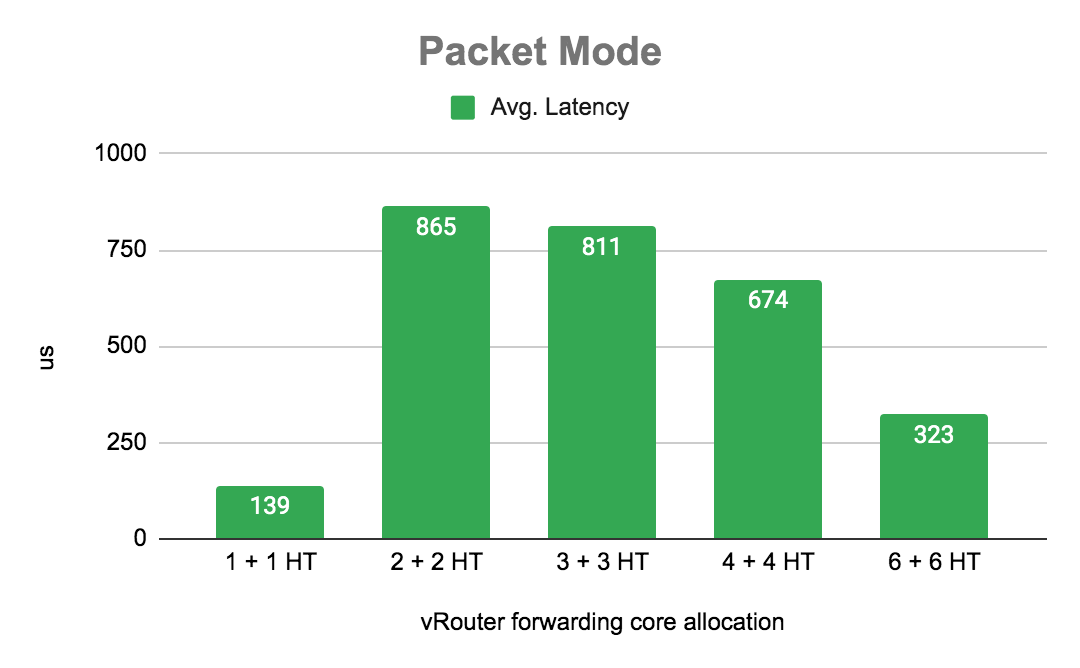
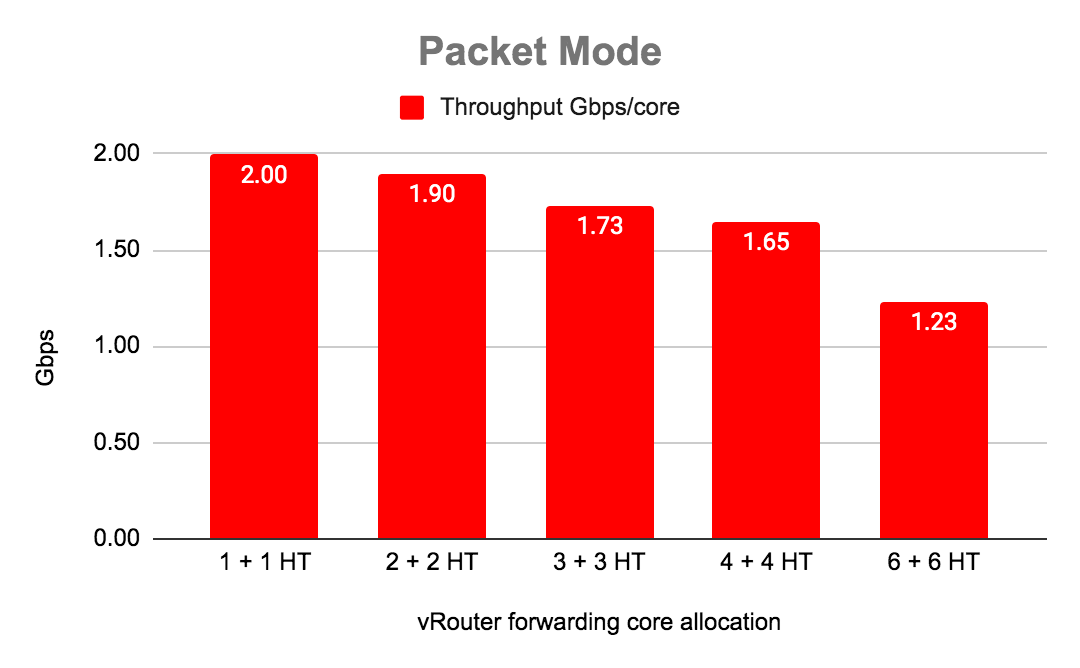


|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Number Cores | Throughput Gbps/core | Total Mpps TX+RX | Mpps/core | Avg. Latency [us] |
| 2 + 0 HT | 1.50 | 4.38 | 2.19 | 116 |
| 3 + 0 HT | 1.40 | 6.31 | 2.10 | 649 |
| 4 + 0 HT | 1.35 | 7.95 | 1.99 | 937 |
| 6 + 0 HT | 1.20 | 10.75 | 1.79 | 946 |

## 

## Packet mode with siblings (HT):

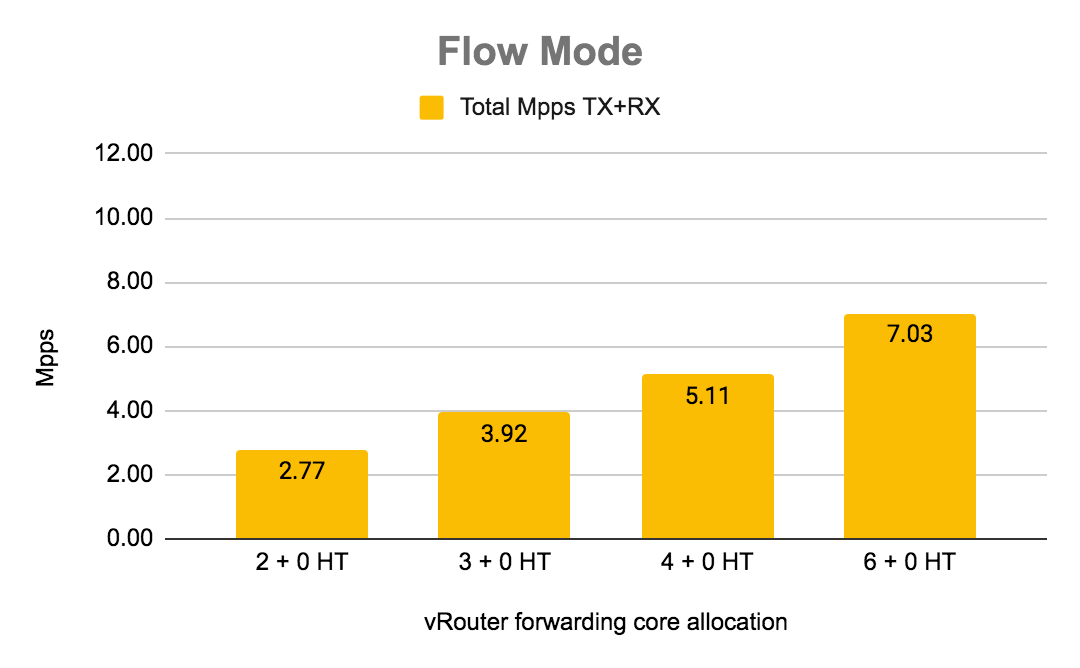
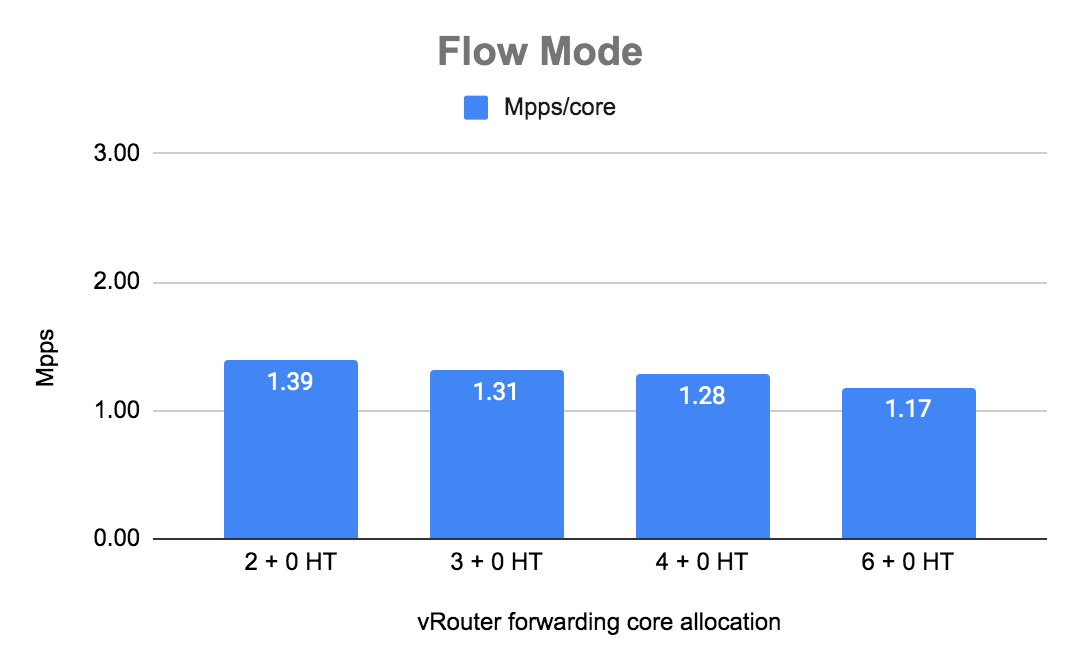


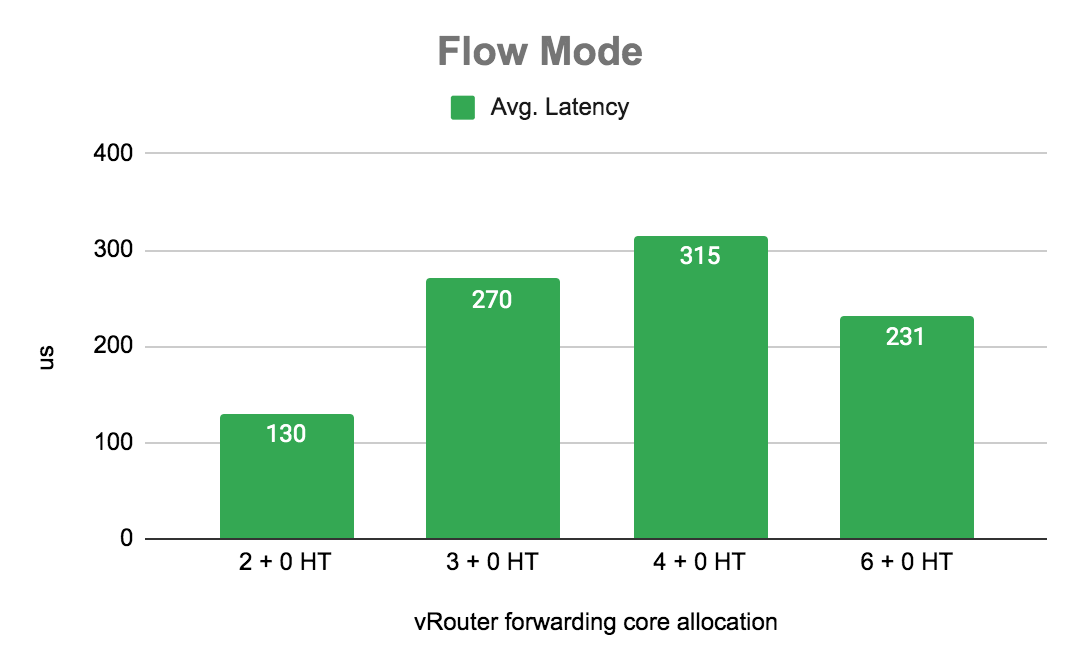
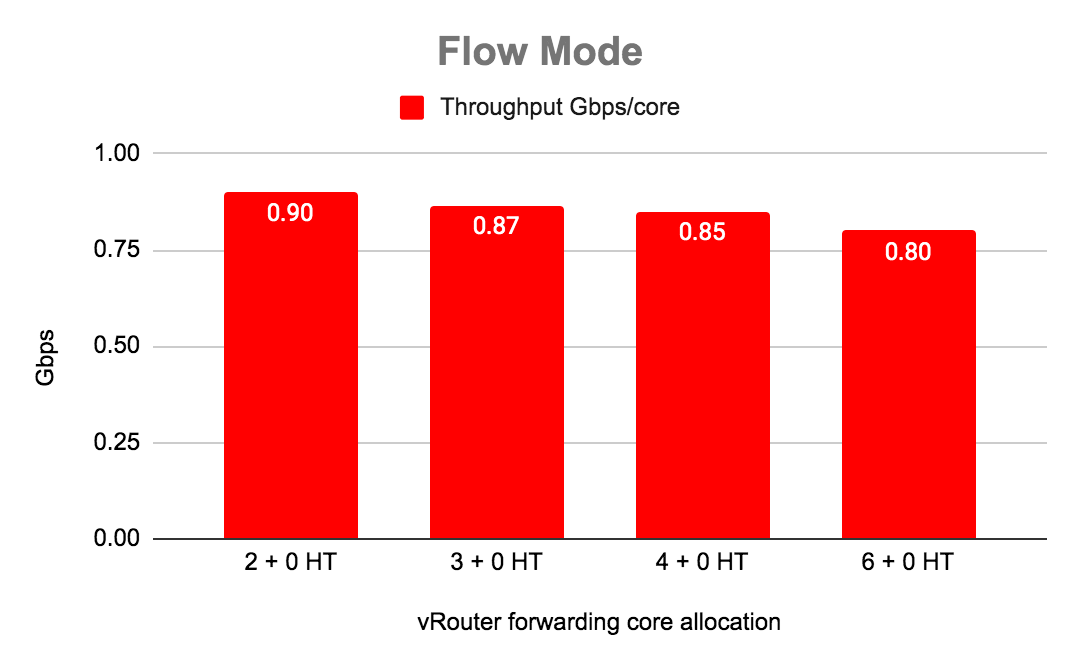


|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Number Cores | Throughput Gbps/core | Total Mpps TX+RX | Mpps/core | Avg. Latency [us] |
| 1 + 1 HT | 2.00 | 3.12 | 3.12 | 139 |
| 2 + 2 HT | 1.90 | 5.56 | 2.78 | 865 |
| 3 + 3 HT | 1.73 | 7.67 | 2.56 | 811 |
| 4 + 4 HT | 1.65 | 9.85 | 2.46 | 674 |
| 6 + 6 HT | 1.23\* | 11.02\* | 1.84 | 323 |

\* results lower than expected due limits of the traffic generator

## Flow mode without siblings (HT):



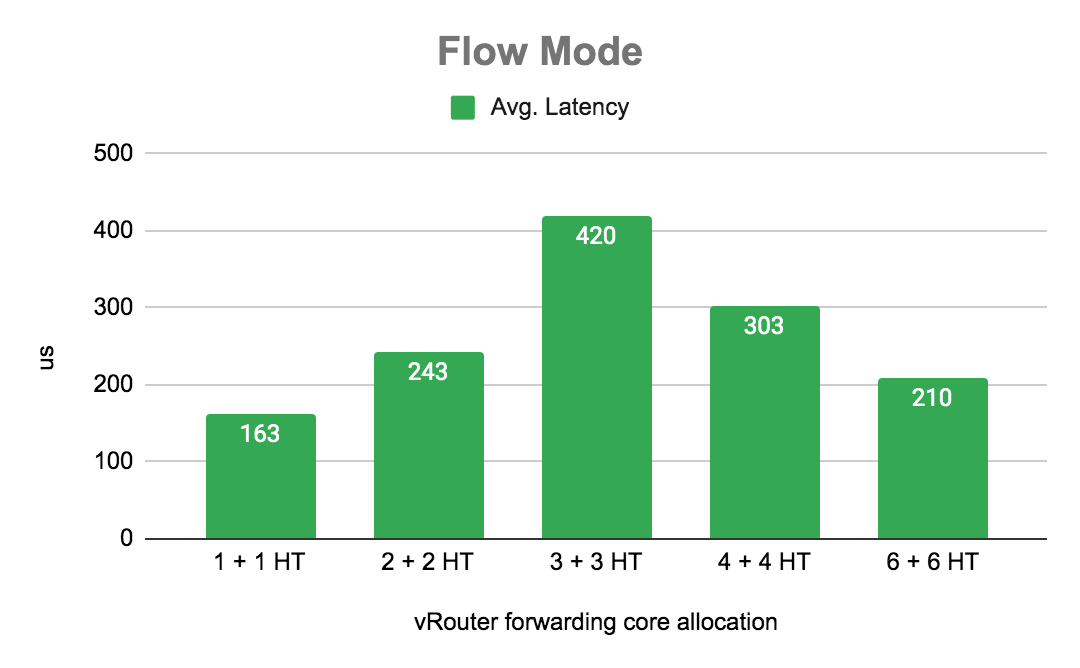
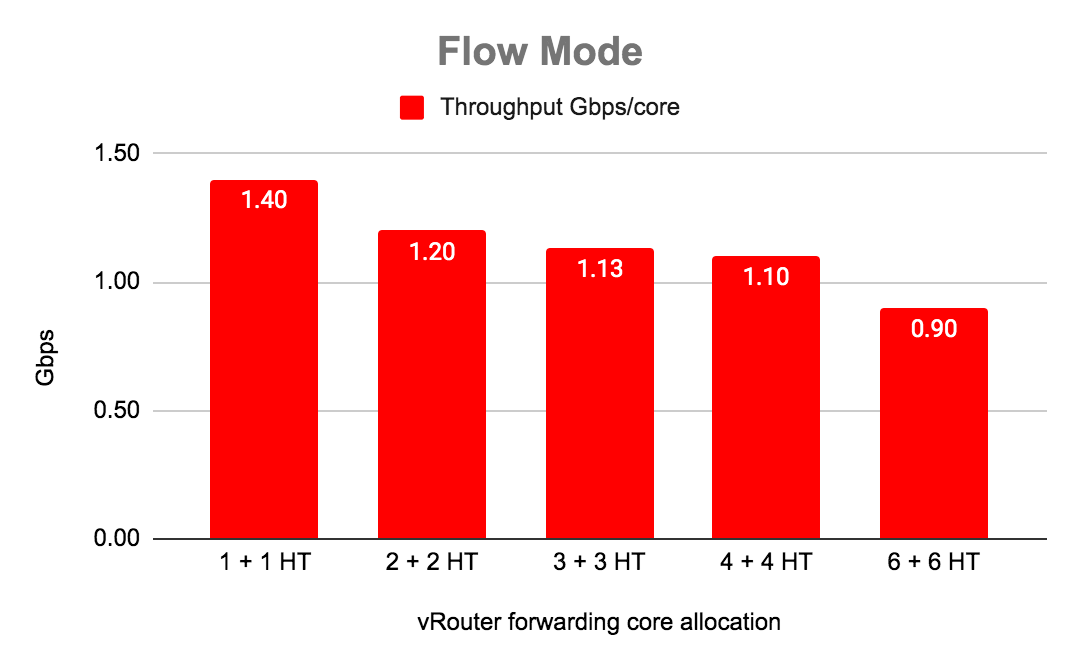
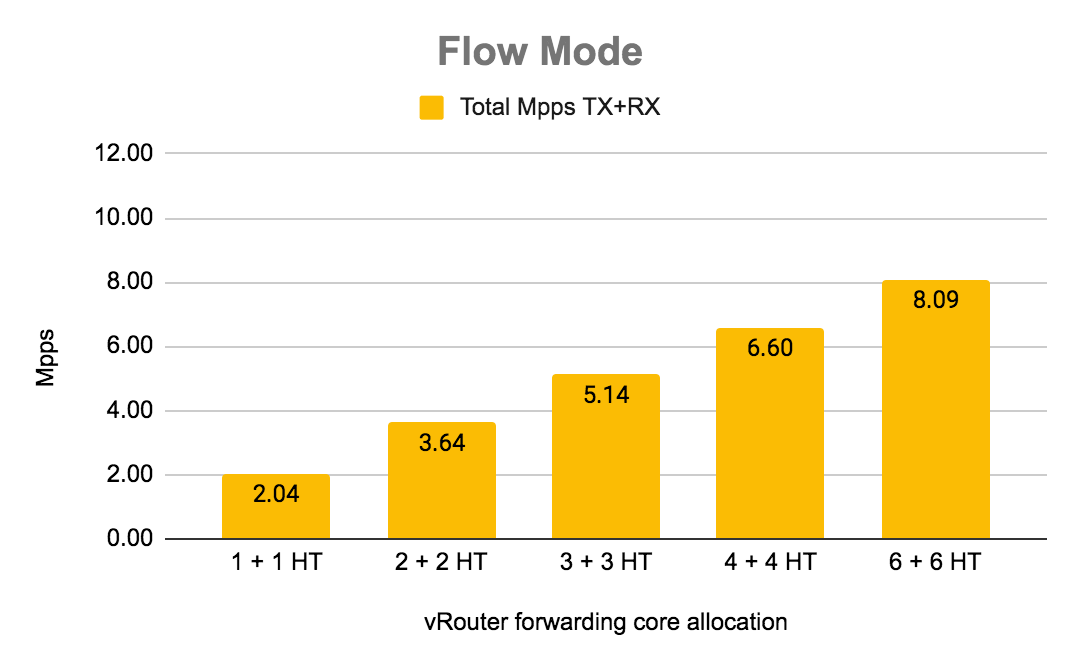
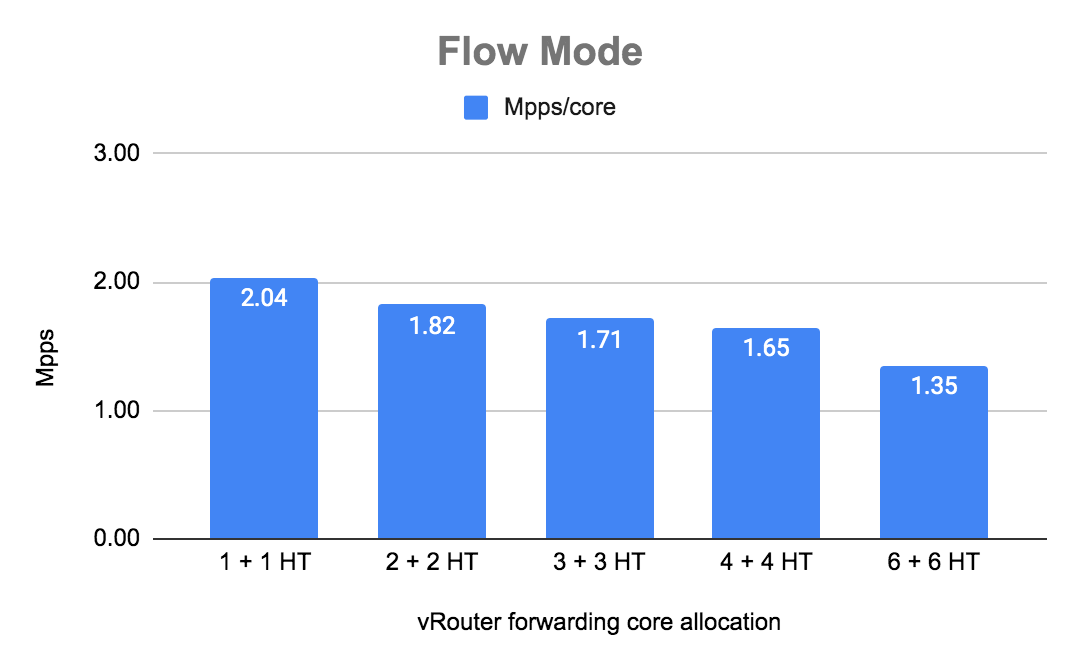


|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Number Cores | Throughput Gbps/core | Total Mpps TX+RX | Mpps/core | Avg. Latency |
| 2 + 0 HT | 0.90 | 2.77 | 1.39 | 130 |
| 3 + 0 HT | 0.87 | 3.92 | 1.31 | 270 |
| 4 + 0 HT | 0.85 | 5.11 | 1.28 | 315 |
| 6 + 0 HT | 0.80 | 7.03 | 1.17 | 231 |

## 

## 

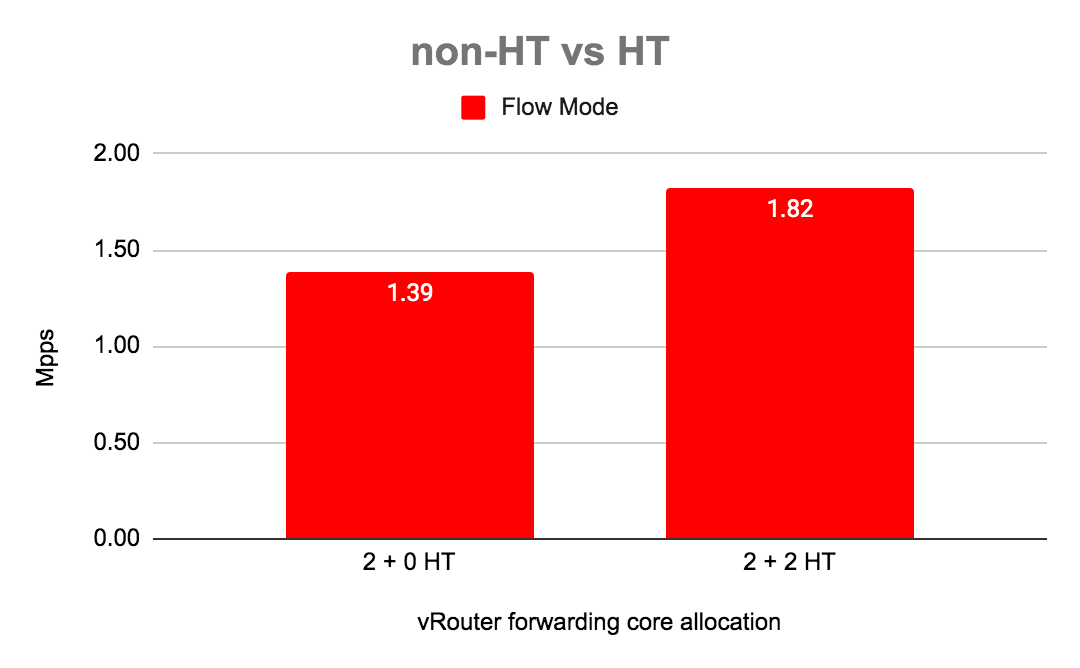
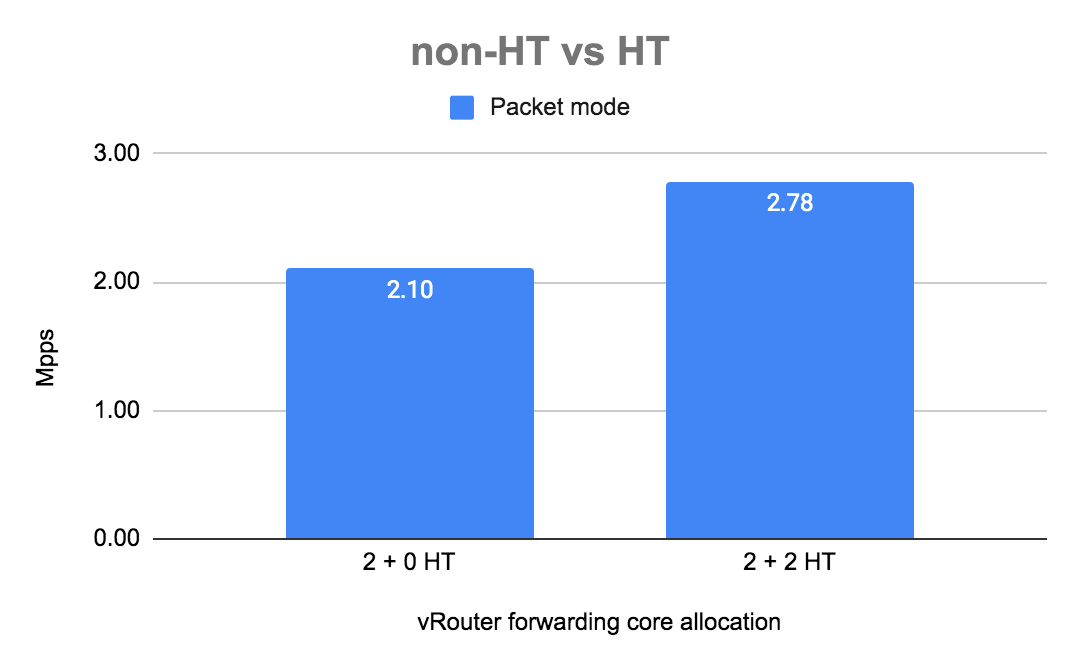
## Flow mode with siblings (HT):



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Number Cores | Throughput Gbps/core | Total Mpps TX+RX | Mpps/core | Avg. Latency |
| 1 + 1 HT | 1.40 | 2.04 | 2.04 | 163 |
| 2 + 2 HT | 1.20 | 3.64 | 1.82 | 243 |
| 3 + 3 HT | 1.13 | 5.14 | 1.71 | 420 |
| 4 + 4 HT | 1.10 | 6.60 | 1.65 | 303 |
| 6 + 6 HT | 0.90 | 8.09 | 1.35 | 210 |

*Summary*

Using HT siblings gives us ~25% of gain as is presented on the diagram below. Enabling HT siblings give as well more queues exposed to VM (VNF) and in consequence better utilisation of VNF forwarding cores. We strongly recommend to use HT siblings for vRouter to improve the performance.



# 

# Test case 6: Number of packets per second (Mpps) per core with increasing number of flows

## Configuration of vRouter on compute-B (DUT):

|  |  |
| --- | --- |
| Operating System | RedHat 7.7 |
| Number of Forwarding Cores | 2 (Physical) + 2 (Siblings) |
| Drop Rate | 0.001% |
| Bond | 2 x 10G (802.3ad LACP) with L3/L4 hash |
| Hugepages | 1G |
| NUMA allocation | vRouter, NIC and VM on the same socket (NUMA 0) |

## CPU allocation (Full-CPU partitioning scheme) on compute-B:

|  |  |
| --- | --- |
| vRouter forwarding cores | 2,4,30,32 |
| Service cores | 0,1,28,29 |
| Control cores | 0,1,28,29 |
| Nova (VM) cores [VM core: HV core] | 0: 18,1: 46, 2: 40, 3: 12, 4: 48, 5: 20, 6: 24, 7: 52, 8: 10, 9: 38 |
| Host OS cores | 0,1,28,29 |
| Kernel isolcpus and tuned isolcpu list | 2-27,30-55 |

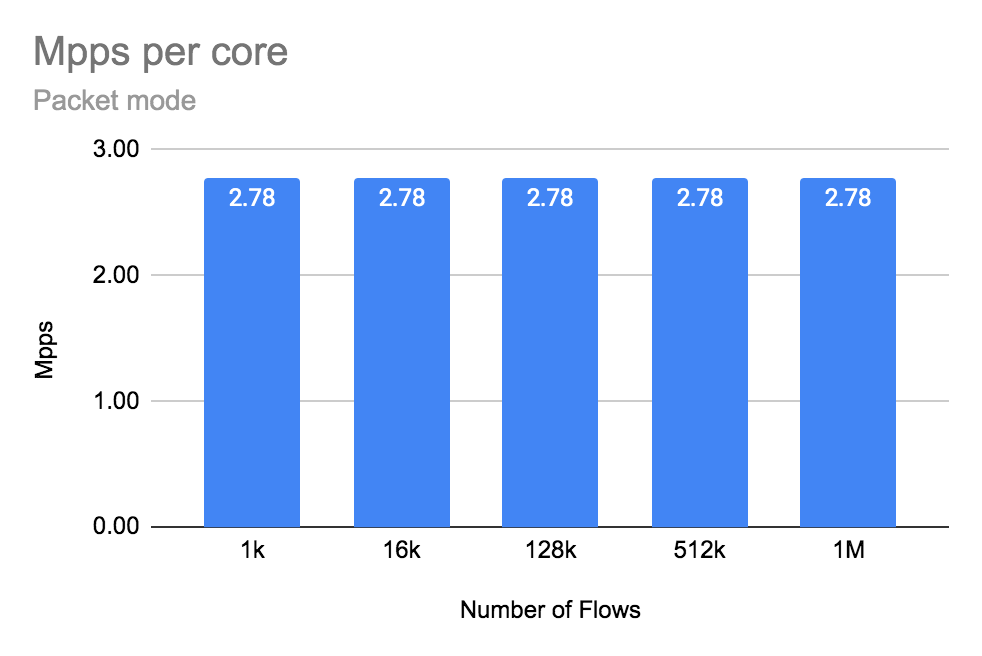
## 

## Traffic pattern

|  |  |
| --- | --- |
| Number of flows | various |
| Encapsulation | MPLSoUDP |
| Packet size | 64B |

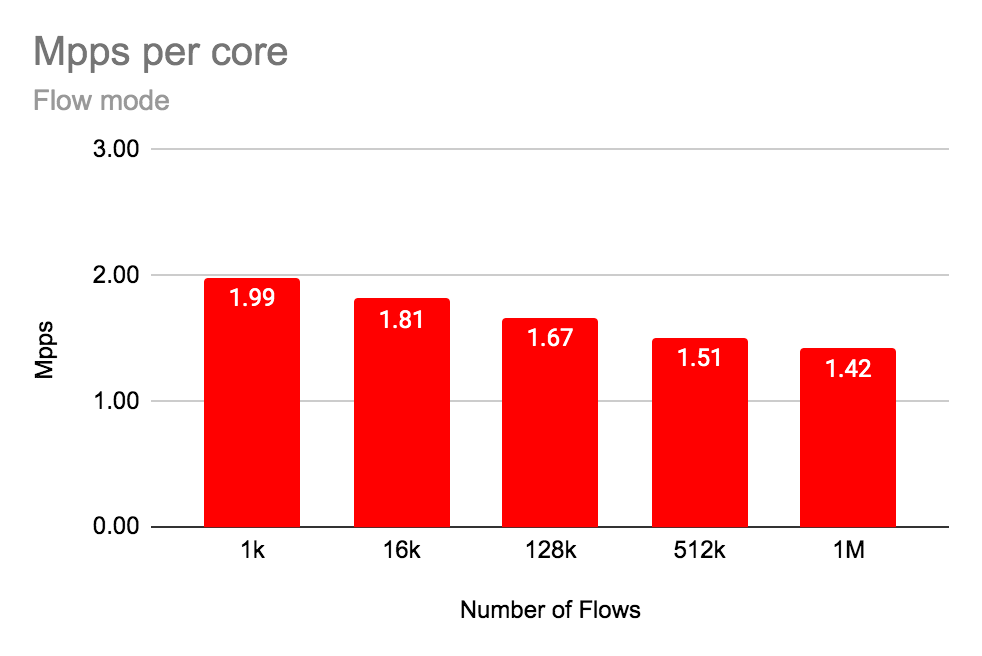
## 

## Packet mode:



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Number of flows | | | | |
|  | 1k | 16k | 128k | 512k | 1M |
| Mpps/core | 2.78 | 2.78 | 2.78 | 2.78 | 2.78 |
| Avg. Latency [us] | 968 | 992 | 970 | 898 | 958 |

## Flow mode:



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Number of flows | | | | |
|  | 1k | 16k | 128k | 512k | 1M |
| Mpps/core | 1.99 | 1.81 | 1.67 | 1.51 | 1.42 |
| Avg. Latency [us] | 976 | 276 | 197 | 178 | 179 |

*Summary*

In packet mode increasing the number of flows does not affect vrouter performance.

Better latency in flow mode is due to lower Mpps (Increases total/per core packet per second increasing latency).

## Intel x710 Fortville configuration

The NIC specific configuration parameters

|  |
| --- |
| cat /etc/sysconfig/network-scripts/ifcfg-vhost0  [...]  BIND\_INT=0000:19:00.0,0000:19:00.1  DRIVER=vfio-pci  BOND\_MODE=4  BOND\_POLICY=layer3+4  CPU\_LIST=2,4,30,32  SERVICE\_CORE\_MASK=0,1,28,29  DPDK\_CTRL\_THREAD\_MASK=0,1,28,29  sed -e "s/\x00/ /g" /proc/$(pidof contrail-vrouter-dpdk)/cmdline  /usr/bin/contrail-vrouter-dpdk --no-daemon --vr\_flow\_entries=2500000 --vr\_mempool\_sz 262144 --dpdk\_ctrl\_thread\_mask 0 --vr\_dpdk\_rx\_ring\_sz 2048 --vr\_dpdk\_tx\_ring\_sz 2048 --yield\_option 0 --service\_core\_mask (0,1,28,29) --dpdk\_ctrl\_thread\_mask (0,1,28,29) --socket-mem 1024 1024 --vdev eth\_bond\_bond1,mode=4,xmit\_policy=l34,socket\_id=0,mac=e4:43:4b:6e:70:b0,lacp\_rate=1,slave=0000:19:00.0,slave=0000:19:00.1 |

# Test case 7: Packets per second (Mpps) with different NUMA placement for VM

## Configuration of vRouter on compute-B (DUT):

|  |  |
| --- | --- |
| Operating System | RedHat 7.7 |
| Number of Forwarding Cores | 2 (Physical) + 2 (Siblings) |
| Drop Rate | 0.001% |
| Bond | 2 x 10G (802.3ad LACP) with L3/L4 hash |
| Hugepages | 1G |
| NUMA allocation | vRouter, NIC and VM NUMA0 or NUMA1 |

## CPU allocation (Full-CPU partitioning scheme) on compute-B:

|  |  |
| --- | --- |
| vRouter forwarding cores | 2,4,30,32 |
| Service cores | 0,1,28,29 |
| Control cores | 0,1,28,29 |
| Nova (VM) cores [VM core: HV core] | 0: 18,1: 46, 2: 40, 3: 12, 4: 48, 5: 20, 6: 24, 7: 52, 8: 10, 9: 38 |
| Host OS cores | 0,1,28,29 |
| Kernel isolcpus and tuned isolcpu list | 2-27,30-55 |

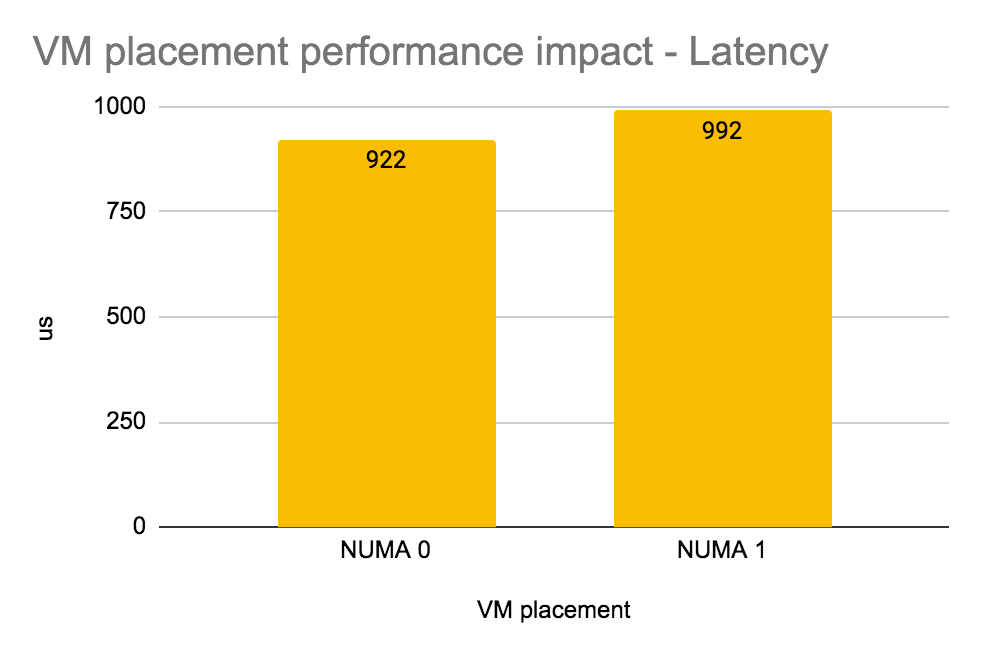
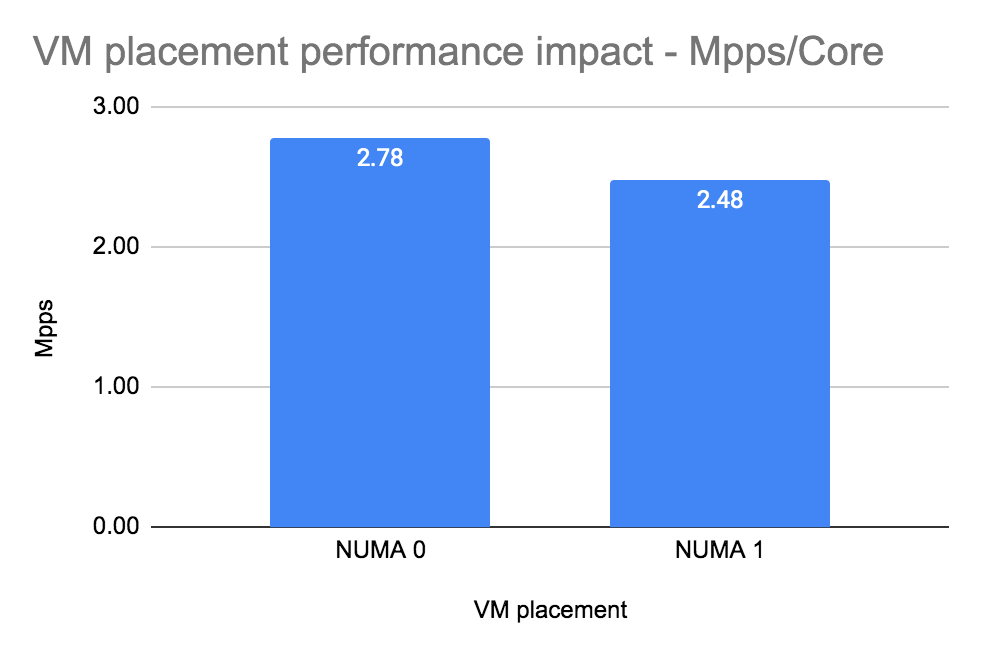
## 

## Traffic pattern

|  |  |
| --- | --- |
| Number of flows | 16535 |
| Encapsulation | MPLSoUDP |
| Packet size | 64B |

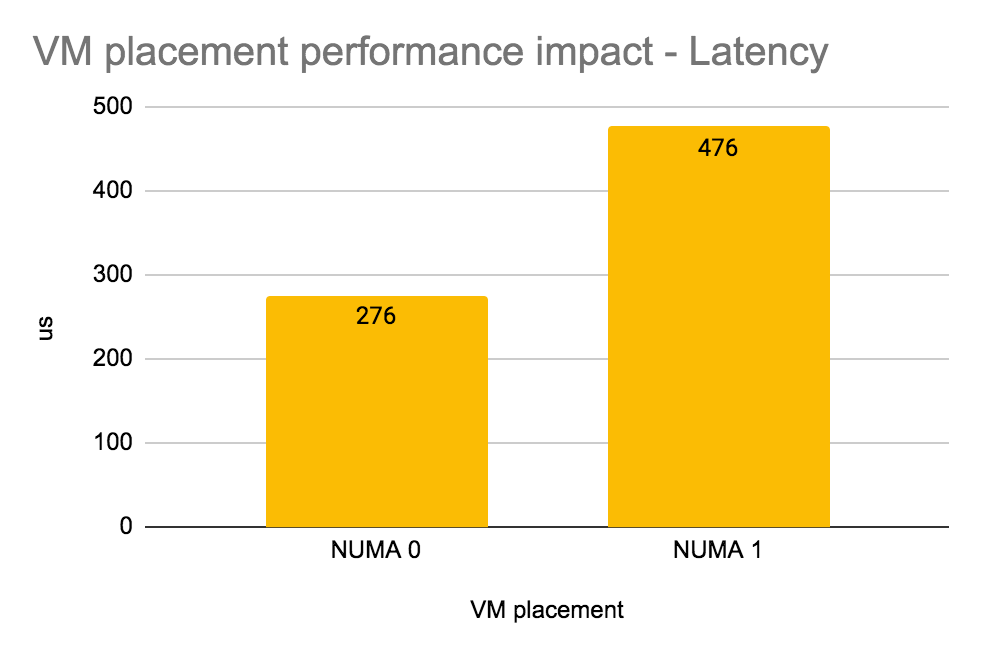
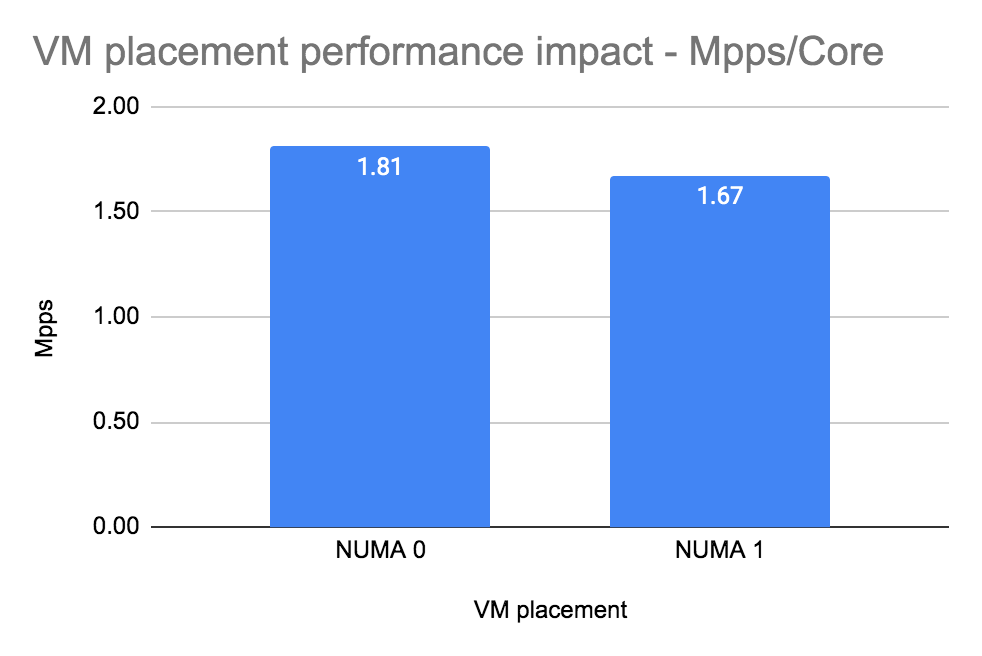
## 

## Packet mode:



|  |  |  |
| --- | --- | --- |
|  | VM placement | |
| NUMA 0 | NUMA 1 |
| Mpps/core | 2.78 | 2.48 |
|
| Avg. Latency [us] | 922 | 992 |
|

## Flow mode:



|  |  |  |
| --- | --- | --- |
|  | VM placement | |
| NUMA 0 | NUMA 1 |
| Mpps/core | 1.81 | 1.67 |
|
| Avg. Latency [us] | 276 | 476 |
|

## 

## Summary

Once VM (VNF) is on second NUMA overall performance decreases ~10% but latency increases especially in flow mode ~ 40%. In packet mode latency degradation is not significant but still is high due to the high load of vRouter queues.

## Intel x710 Fortville configuration

The NIC specific configuration parameters

|  |
| --- |
| cat /etc/sysconfig/network-scripts/ifcfg-vhost0  [...]  BIND\_INT=0000:19:00.0,0000:19:00.1  DRIVER=vfio-pci  BOND\_MODE=4  BOND\_POLICY=layer3+4  CPU\_LIST=2,4,30,32  SERVICE\_CORE\_MASK=0,1,28,29  DPDK\_CTRL\_THREAD\_MASK=0,1,28,29  sed -e "s/\x00/ /g" /proc/$(pidof contrail-vrouter-dpdk)/cmdline  /usr/bin/contrail-vrouter-dpdk --no-daemon --vr\_flow\_entries=2500000 --vr\_mempool\_sz 262144 --dpdk\_ctrl\_thread\_mask 0 --vr\_dpdk\_rx\_ring\_sz 2048 --vr\_dpdk\_tx\_ring\_sz 2048 --yield\_option 0 --service\_core\_mask (0,1,28,29) --dpdk\_ctrl\_thread\_mask (0,1,28,29) --socket-mem 1024 1024 --vdev eth\_bond\_bond1,mode=4,xmit\_policy=l34,socket\_id=0,mac=e4:43:4b:6e:70:b0,lacp\_rate=1,slave=0000:19:00.0,slave=0000:19:00.1 |

# 

# Test case 8: Number of packets per second (Mpps) per core with different HugePage sizes

## Configuration of vRouter on compute-B (DUT):

|  |  |
| --- | --- |
| Operating System | RedHat 7.7 |
| Number of Forwarding Cores | 2 (Physical) + 2 (Siblings) |
| Drop Rate | 0.001% |
| Bond | 2 x 10G (802.3ad LACP) with L3/L4 hash |
| Hugepages | 2M or 1G |
| NUMA allocation | vRouter, NIC and VM NUMA0 or NUMA1 |

## CPU allocation (Full-CPU partitioning scheme) on compute-B:

|  |  |
| --- | --- |
| vRouter forwarding cores | 2,4,30,32 |
| Service cores | 0,1,28,29 |
| Control cores | 0,1,28,29 |
| Nova (VM) cores [VM core: HV core] | 0: 18,1: 46, 2: 40, 3: 12, 4: 48, 5: 20, 6: 24, 7: 52, 8: 10, 9: 38 |
| Host OS cores | 0,1,28,29 |
| Kernel isolcpus and tuned isolcpu list | 2-27,30-55 |

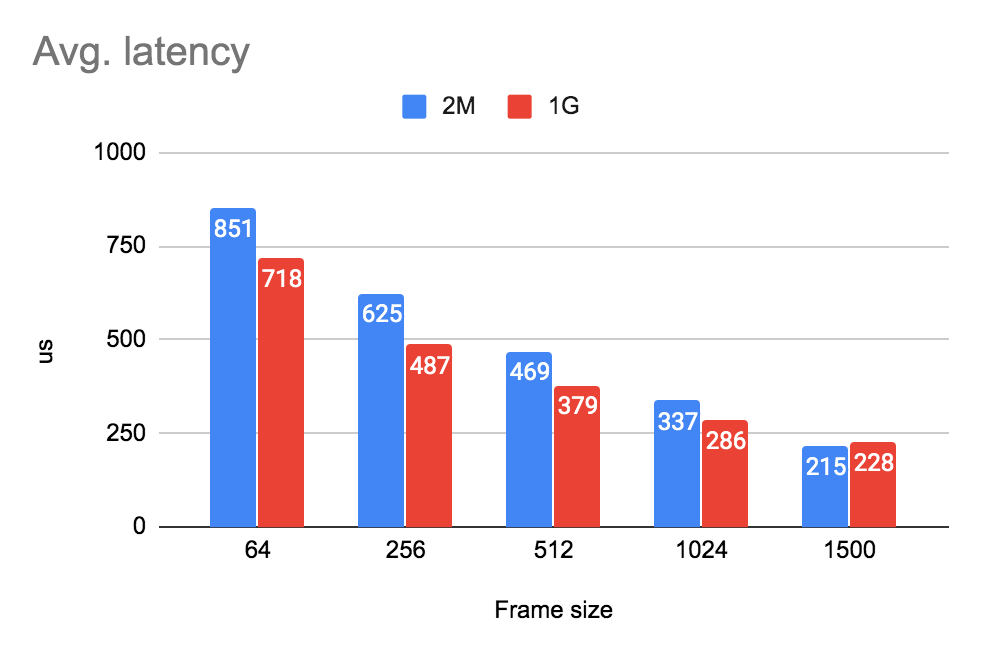
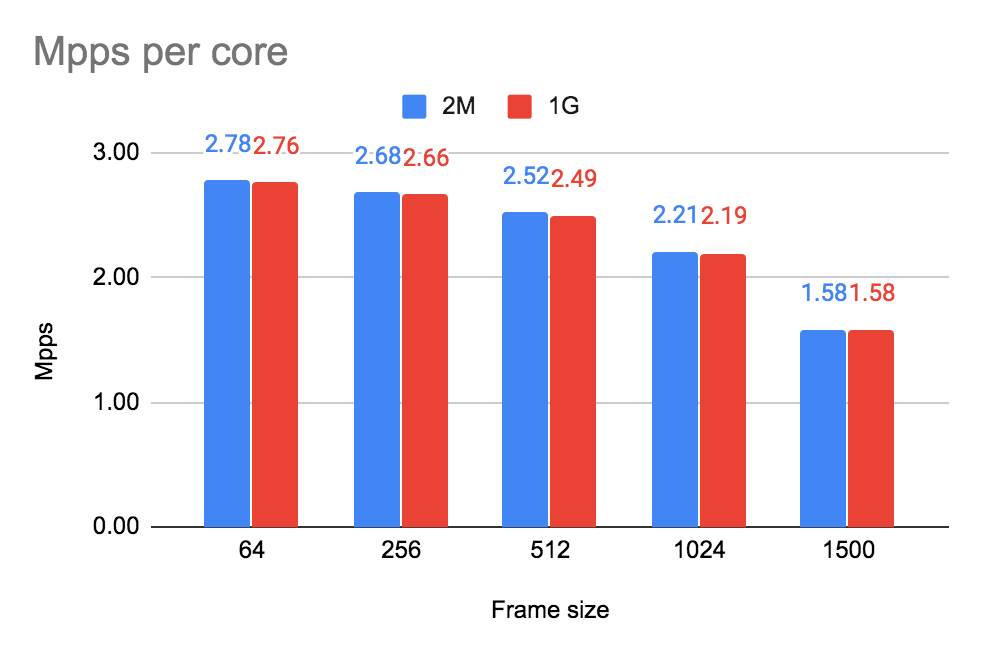
## 

## Traffic pattern

|  |  |
| --- | --- |
| Number of flows | 16535 |
| Encapsulation | MPLSoUDP |
| Packet size | 64B |

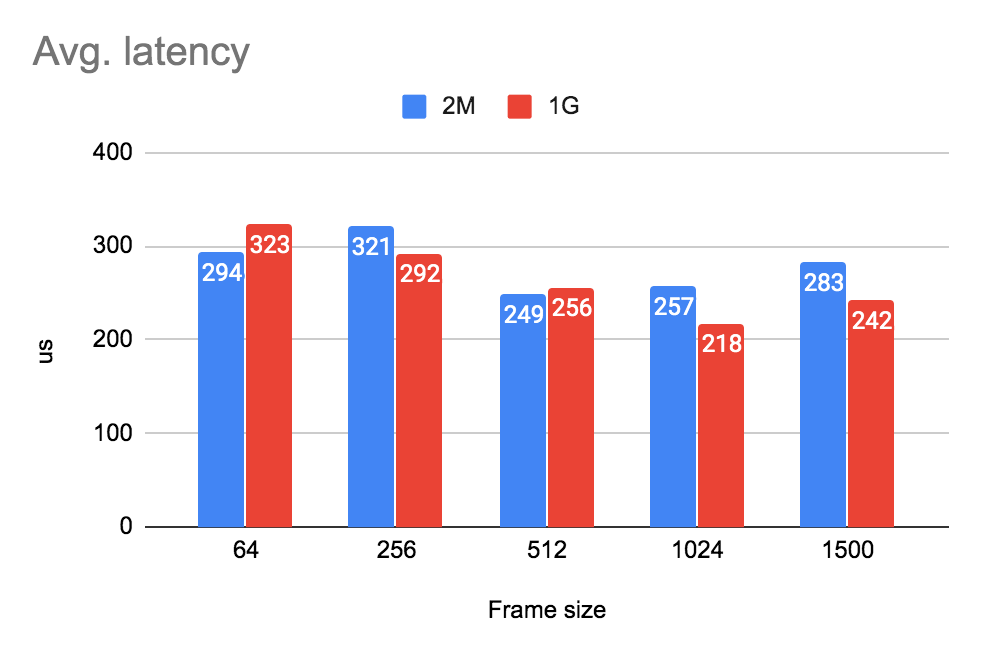
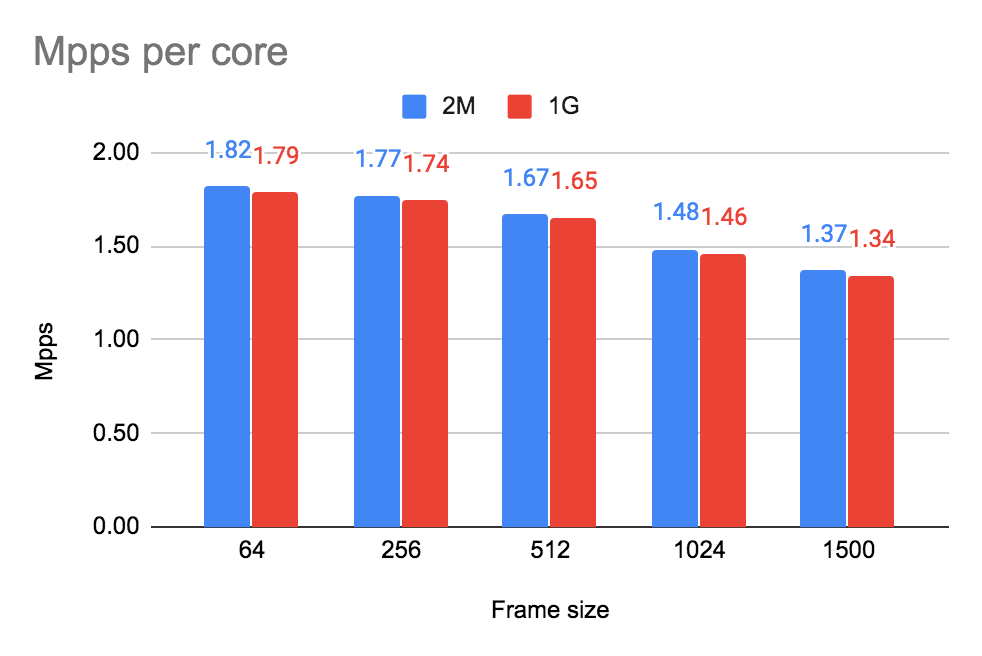
## 

## Packet mode:



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Huge Page Size | Frame size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| Mpps/core | 2M | 2.78 | 2.68 | 2.52 | 2.21 | 1.58 |
| 1G | 2.76 | 2.66 | 2.49 | 2.19 | 1.58 |
| Avg. Latency [us] | 2M | 851 | 625 | 469 | 337 | 215 |
| 1G | 718 | 487 | 379 | 286 | 228 |

## Flow mode:



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Huge Page Size | Frame size [bytes] | | | | |
| 64 | 256 | 512 | 1024 | 1500 |
| Mpps/core | 2M | 1.82 | 1.77 | 1.67 | 1.48 | 1.37 |
| 1G | 1.79 | 1.74 | 1.65 | 1.46 | 1.34 |
| Avg. Latency [us] | 2M | 294 | 321 | 249 | 257 | 283 |
| 1G | 323 | 292 | 256 | 218 | 242 |

## 

## Summary

Usage of 2M instead of default 1G huge pages, increases slightly the performance but degrades latency.

## Intel x710 Fortville configuration

The NIC specific configuration parameters

|  |
| --- |
| cat /etc/sysconfig/network-scripts/ifcfg-vhost0  [...]  BIND\_INT=0000:19:00.0,0000:19:00.1  DRIVER=vfio-pci  BOND\_MODE=4  BOND\_POLICY=layer3+4  CPU\_LIST=2,4,30,32  SERVICE\_CORE\_MASK=0,1,28,29  DPDK\_CTRL\_THREAD\_MASK=0,1,28,29  sed -e "s/\x00/ /g" /proc/$(pidof contrail-vrouter-dpdk)/cmdline  /usr/bin/contrail-vrouter-dpdk --no-daemon --vr\_flow\_entries=2500000 --vr\_mempool\_sz 262144 --dpdk\_ctrl\_thread\_mask 0 --vr\_dpdk\_rx\_ring\_sz 2048 --vr\_dpdk\_tx\_ring\_sz 2048 --yield\_option 0 --service\_core\_mask (0,1,28,29) --dpdk\_ctrl\_thread\_mask (0,1,28,29) --socket-mem 1024 1024 --vdev eth\_bond\_bond1,mode=4,xmit\_policy=l34,socket\_id=0,mac=e4:43:4b:6e:70:b0,lacp\_rate=1,slave=0000:19:00.0,slave=0000:19:00.1 |

# 

# Appendix

## Guide how to run testing framework

Deployment of Rapid on Contrail environment using publicly available repository of heat templates <https://github.com/damjanek/dpdk-prox-contrail> (simple version without ansible automation)

Deployment of Rapid on Contrail environment using Juniper internalrepository of heat templates <https://ssd-git.juniper.net/sre/dpdk-testing> (advance version with ansible automation - optionally)

Rapid image gdrive (Juniper only) <https://drive.google.com/file/d/1bX6e7RdKzmay4yisaiqYcXT-a4B0EUCN/view?usp=sharing>

## Configure high performance Contrail Cloud computes

For compute nodes which will be acting as a traffic generator (gen VM) and looping (swap VM) please use the following template snippets.

site.yml

|  |
| --- |
| overcloud:  extra\_config:  NovaSchedulerDefaultFilters:  - RetryFilter  - DifferentHostFilter  - SameHostFilter  - AvailabilityZoneFilter  - ComputeFilter  - ComputeCapabilitiesFilter  - ImagePropertiesFilter  - ServerGroupAntiAffinityFilter  - ServerGroupAffinityFilter  - NUMATopologyFilter  - AggregateInstanceExtraSpecsFilter  NovaComputeExtraConfig:  nova::cpu\_allocation\_ratio: 1.0  nova::ram\_allocation\_ratio: 1.0  nova::disk\_allocation\_ratio: 1.0  #  # node 0 cpus: 0 2 4 6 8 10 12 14 16 18 20 22 24 26  # 28 30 32 34 36 38 40 42 44 46 48 50 52 54  #  # node 1 cpus: 1 3 5 7 9 11 13 15 17 19 21 23 25 27  # 29 31 33 35 37 39 41 43 45 47 49 51 53 55  #  # OS: 0,1,28,29 (4 cores, including siblings)  # vRouter: 2,4,6,8,30,32,34,36  # Nova: rest of available cores from NUMA0 and optionally NUMA1  # Compute Node for Swap VM (min CPU 2+2HT good to have ability to increase during tests)  # Compute Node for the Generator (min CPU 4+4HT)    ComputeDpdkParameters:  TunedProfileName: "cpu-partitioning"  ContrailDpdkOptions: "--vr\_flow\_entries=2500000 --vr\_mempool\_sz 262144 --vr\_dpdk\_rx\_ring\_sz 2048 --vr\_dpdk\_tx\_ring\_sz 2048 --yield\_option 0"  ContrailVrouterHugepages1GB: 220  ContrailVrouterHugepages2MB: 8192  IsolCpusList: "2-27,30-55"  KernelArgs: "isolcpus=2-27,30-55"  NovaVcpuPinSet: ['10','12','14','16','18','20','22','24','26','38','40','42','44','46','48','50','52','54']  nova::cpu\_allocation\_ratio: 1.0  nova::ram\_allocation\_ratio: 1.0  nova::disk\_allocation\_ratio: 1.0  ContrailSettings:  SERVICE\_CORE\_MASK: '0,1,28,29'  DPDK\_CTRL\_THREAD\_MASK: '0,1,28,29'  LACP\_RATE: 1 |

overcloud-nics.yml

|  |
| --- |
| # Compute Node for Swap VM (min CPU 2+2HT good to have possibility to increase during tests)  # Compute Node for the Generator (min CPU 4+4HT)  ComputeDpdk\_network\_config:  - type: interface  name: p1p1  dns\_servers:  get\_param: DnsServers  use\_dhcp: false  mtu:  get\_param: ControlPlaneNetworkMtu  addresses:  - ip\_netmask:  list\_join:  - '/'  - - get\_param: ControlPlaneIp  - get\_param: ControlPlaneSubnetCidr  routes:  -  ip\_netmask: 169.254.169.254/32  next\_hop:  get\_param: EC2MetadataIp  -  default: True  next\_hop:  get\_param: ControlPlaneDefaultRoute  use\_dhcp: false  - type: linux\_bond  name: ten2  use\_dhcp: false  bonding\_options: "mode=802.3ad xmit\_hash\_policy=layer3+4 lacp\_rate=fast updelay=1000 miimon=100"  mtu:  get\_param: InternalApi1NetworkMtu  members:  - type: interface  name: em3  primary: true  mtu:  get\_param: InternalApi1NetworkMtu  - type: interface  name: em4  mtu:  get\_param: InternalApi1NetworkMtu  - type: vlan  device: ten2  vlan\_id:  get\_param: InternalApi1NetworkVlanID  mtu:  get\_param: InternalApi1NetworkMtu  addresses:  - ip\_netmask:  get\_param: InternalApi1IpSubnet  routes:  -  ip\_netmask:  get\_param: InternalApiSupernet  next\_hop:  get\_param: InternalApi1InterfaceDefaultRoute  - type: vlan  device: ten2  vlan\_id:  get\_param: Storage1NetworkVlanID  mtu:  get\_param: Storage1NetworkMtu  addresses:  - ip\_netmask:  get\_param: Storage1IpSubnet  routes:  -  ip\_netmask:  get\_param: StorageSupernet  next\_hop:  get\_param: Storage1InterfaceDefaultRoute  - type: contrail\_vrouter\_dpdk  name: vhost0  driver: "{{ overcloud['contrail']['vrouter']['dpdk']['driver'] }}"  bond\_mode: 4  bond\_policy: layer3+4  **cpu\_list: "2,4,30,32"**  **# For the Generator set cpu\_list: "2,4,6,8,30,32,34,36"**  members:  - type: interface  name: em1  - type: interface  name: em2  mtu:  get\_param: Tenant1NetworkMtu  addresses:  - ip\_netmask:  get\_param: Tenant1IpSubnet  routes:  -  ip\_netmask:  get\_param: TenantSupernet  next\_hop:  get\_param: Tenant1InterfaceDefaultRoute |

## Configure high performance Contrail Networking on computes with TripleO heat templates

For compute nodes which will be acting as a traffic generator (gen VM) and looping (swap VM) please use the following template snippets.

environments/contrail/environment-extra.yaml

|  |
| --- |
| ComputeDpdk1Hw0Parameters:  ContrailDpdkOptions: --vr\_flow\_entries=2500000 --vr\_mempool\_sz 262144 --vr\_dpdk\_rx\_ring\_sz 2048 --vr\_dpdk\_tx\_ring\_sz 2048 --yield\_option 0  ContrailSettings:  DPDK\_CTRL\_THREAD\_MASK: '0,1,28,29'  LACP\_RATE: 1  SERVICE\_CORE\_MASK: '0,1,28,29'  ContrailVrouterHugepages1GB: 220  ContrailVrouterHugepages2MB: 8192  IsolCpusList: 2-27,30-55  KernelArgs: isolcpus=2-27,30-55  NovaVcpuPinSet:  - '10'  - '12'  - '14'  - '16'  - '18'  - '20'  - '22'  - '24'  - '26'  - '38'  - '40'  - '42'  - '44'  - '46'  - '48'  - '50'  - '52'  - '54'  TunedProfileName: cpu-partitioning  nova::cpu\_allocation\_ratio: 1.0  nova::disk\_allocation\_ratio: 1.0  nova::ram\_allocation\_ratio: 1.0 |

overcloud-nics.yml

|  |
| --- |
| # Compute Node for Swap VM (min CPU 2+2HT good to have possibility to increase during tests)  # Compute Node for the Generator (min CPU 4+4HT)  ComputeDpdk\_network\_config:  - type: interface  name: p1p1  dns\_servers:  get\_param: DnsServers  use\_dhcp: false  mtu:  get\_param: ControlPlaneNetworkMtu  addresses:  - ip\_netmask:  list\_join:  - '/'  - - get\_param: ControlPlaneIp  - get\_param: ControlPlaneSubnetCidr  routes:  -  ip\_netmask: 169.254.169.254/32  next\_hop:  get\_param: EC2MetadataIp  -  default: True  next\_hop:  get\_param: ControlPlaneDefaultRoute  use\_dhcp: false  - type: linux\_bond  name: ten2  use\_dhcp: false  bonding\_options: "mode=802.3ad xmit\_hash\_policy=layer3+4 lacp\_rate=fast updelay=1000 miimon=100"  mtu:  get\_param: InternalApi1NetworkMtu  members:  - type: interface  name: em3  primary: true  mtu:  get\_param: InternalApi1NetworkMtu  - type: interface  name: em4  mtu:  get\_param: InternalApi1NetworkMtu  - type: vlan  device: ten2  vlan\_id:  get\_param: InternalApi1NetworkVlanID  mtu:  get\_param: InternalApi1NetworkMtu  addresses:  - ip\_netmask:  get\_param: InternalApi1IpSubnet  routes:  -  ip\_netmask:  get\_param: InternalApiSupernet  next\_hop:  get\_param: InternalApi1InterfaceDefaultRoute  - type: vlan  device: ten2  vlan\_id:  get\_param: Storage1NetworkVlanID  mtu:  get\_param: Storage1NetworkMtu  addresses:  - ip\_netmask:  get\_param: Storage1IpSubnet  routes:  -  ip\_netmask:  get\_param: StorageSupernet  next\_hop:  get\_param: Storage1InterfaceDefaultRoute  - type: contrail\_vrouter\_dpdk  name: vhost0  driver: "{{ overcloud['contrail']['vrouter']['dpdk']['driver'] }}"  bond\_mode: 4  bond\_policy: layer3+4  **cpu\_list: "2,4,30,32"**  **# For the Generator set cpu\_list: "2,4,6,8,30,32,34,36"**  members:  - type: interface  name: em1  - type: interface  name: em2  mtu:  get\_param: Tenant1NetworkMtu  addresses:  - ip\_netmask:  get\_param: Tenant1IpSubnet  routes:  -  ip\_netmask:  get\_param: TenantSupernet  next\_hop:  get\_param: Tenant1InterfaceDefaultRoute |

# 

# 

# References

Raw data document

<https://docs.google.com/spreadsheets/d/1Od1eaKI_5W_NYszRNxqUKTLWRuqXS6OhGHjjI1wBO1Q/edit#gid=1753042829>

Rapid wiki <https://wiki.opnfv.org/display/SAM/Rapid+scripting>