ETSI GS NFV-IFA 003 V2.3.1 (2017-08)

Network Functions Virtualisation (NFV) Release 2; Acceleration Technologies; vSwitch Benchmarking and Acceleration Specification

Disclaimer

The present document has been produced and approved by the Network Functions Virtualisation (NFV) ETSI Industry Specification Group (ISG) and represents the views of those members who participated in this ISG.
It does not necessarily represent the views of the entire ETSI membership.

Reference
RGS/NFV-IFA003ed231

Keywords
acceleration, benchmarking, NFV, performance, requirements, switching, virtualisation

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

This Group Specification (GS) has been produced by ETSI Industry Specification Group (ISG) Network Functions Virtualisation (NFV).

Modal verbs terminology

In the present document "**shall**", "**shall not**", "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

1 Scope

The present document specifies performance benchmarking metrics for virtual switching, with the goal that the metrics will adequately quantify performance gains achieved through virtual switch acceleration conforming to the associated requirements specified herein. The acceleration-related requirements will be applicable to common virtual switching functions across usage models such as packet delivery into VNFs, network overlay and tunnel termination, stateful Network Address Translators (NAT), service chaining, load balancing and, in general, match-action based policies/flows applied to traffic going to/from the VMs. The present document will also provide deployment scenarios with applicability to multiple vendor implementations and recommendations for follow-on proof of concept activities.

2 References

2.1 Normative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are necessary for the application of the present document.

[1] ETSI GS NFV 003: "Network Functions Virtualisation (NFV); Terminology for Main Concepts in NFV".

[2] IETF RFC 2544: "Benchmarking Methodology for Network Interconnect Devices".

[3] IETF RFC 7679: "A One-Way Delay Metric for IP Performance Metrics (IPPM)".

[4] IETF RFC 7680: "A One-Way Loss Metric for IP Performance Metrics (IPPM)".

[5] IETF RFC 3511: "Benchmarking Methodology for Firewall Performance".

[6] IETF RFC 4737: "Packet Reordering Metrics".

[7] IETF RFC 5481: "Packet Delay Variation Applicability Statement".

[8] IETF RFC 6703: "Reporting IP Network Performance Metrics: Different Points of View".

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] IETF draft-ietf-bmwg-ipsec-term-12.txt: "Terminology for Benchmarking IPsec Devices".

[i.2] IETF draft-ietf-bmwg-ipsec-meth-05.txt: "Methodology for Benchmarking IPsec Devices".

[i.3] IETF draft-ietf-bmwg-virtual-net-01.txt: "Considerations for Benchmarking Virtual Network Functions and Their Infrastructure".

[i.4] IETF draft-vsperf-bmwg-vswitch-opnfv-01.txt: "Benchmarking Virtual Switches in OPNFV".

[i.5] IETF RFC 6049: "Spatial Composition of Metrics".

[i.6] IETF RFC 7348: "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks".

[i.7] ETSI GS NFV-INF 007: "Network Functions Virtualisation (NFV); Infrastructure; Methodology to describe Interfaces and Abstractions".

[i.8] ETSI GS NFV-IFA 002: "Network Functions Virtualisation (NFV); Acceleration Technologies; VNF Interfaces Specification".

[i.9] IETF RFC 6815: "Applicability Statement for RFC 2544: Use on Production Networks Considered Harmful".

[i.10] IETF RFC 6985: "IMIX Genome: Specification of Variable Packet Sizes for Additional Testing".

3 Definitions and abbreviations

3.1 Definitions

For the purposes of the present document, the terms and definitions given in ETSI GS NFV 003 [1] apply.

3.2 Abbreviations

For the purposes of the present document, the following abbreviations apply:

ACK	Acknowledge
ACPI	Advanced Configuration and Power Interface
ARP	Address Resolution Protocol
BIOS	Basic Input Output System
BS	Block Storage
CPU	Central Processing Unit
DIMM	Dual In-line Memory Module
DPDK	Data Plane Development Kit
DUT	Device Under Test
GRUB	Grand Unified Bootloader
HTTP	HyperText Transfer Protocol
IMIX	Internet MIX
IO	Input Output
IP	Internet Protocol
MAC	Media Access Control
MTU	Maximum Transmission Unit
NAT	Network Address Translation
NFP	Network Forwarding Path
NFV	Network Functions Virtualisation
NFVI	NFV Infrastructure
NIC	Network Interface Card
NUMA	Non Uniform Memory Access
OPNFV	Open Platform for NFV
PCI	Peripheral Component Interconnect
PDV	Packet Delay Variation
RSS	Receive Side Scaling
SF	Service Function
SFC	Service Function Chaining

SFF	Service Function Forwarders
SLA	Service Level Agreement
SUT	System Under Test
SW	Software
TCP	Transmission Control Protocol
TSO	TCP Segment Offload
VBS	Virtual Block Storage
VIM	Virtual Infrastructure Manager
VM	Virtual Machine
VNF	Virtualised Network Function
VNFC	Virtual Network Function Component
VNFD	Virtual Network Function Description
VNFFG	VNF Forwarding Graph
VNI	VxLAN Network Identifier
VSPERF	OPNFV vSwitch Performance Project
VXLAN	Virtual eXtensible Local Area Network

4 Overview

4.1 Problem Statement

Inside a compute platform a virtual switch (vSwitch) is used to interconnect VNFs that share the same platform. The vSwitch is in a unique position of being at the intersection of the network and the VNFs themselves. As such the implementation and specifics of the virtual switching on the platform need to be transparent to the VNFs in the system, thus as opposed to ETSI GS NFV-IFA 002 [i.8] where VNFs are requesting acceleration, vSwitch acceleration needs to be transparent to individual VNFs, and controlled from the VIM.

NOTE: In the context of the present document, vSwitch may include some of the functionality of vRouter as defined in ETSI GS NFV-INF 007 [i.7].

A 'flow' within the vSwitch is given as the classification (locator + domain) and the port forwarding action. This definition of flow differs from what is tracked as a flow within a VNF or even within the VIM. In this context the vSwitch is only concerned with the flow information needed to perform the virtual switching functionality.

Since the vSwitch finds itself in this unique position between the rest of the network and the VNFs, it is very common to add additional functionality at this point of control. In the present document, these are called 'in-line functions', and are defined as network services that have been placed in-line with the switching function. Examples of in-line functions include:

- ACLs: Doing a more complex (usually wildcard based) classification for security and monitoring purposes
- Tunnel Endpoint: Pushing packets in and out of a tunnel in order to traverse a physical network
- Address Translation/NAT: Translating packet headers to expand the address space of the network
- Load Balancing: Choosing from a set of destinations to forward a packet
- QoS: assigning a class of service for the purpose of traffic shaping, rate limiting, priority queuing, mapping of per-packet features of VNF to infrastructure

These in-line functions are logically separate from the baseline virtual switching function, and as such may have their own specific definition of what constitutes a 'flow', and what additional classification and state information is tracked.

In addition to above in-line functions, stateful operations such as Firewall or Load Balancer may be implemented.

In contrast to in-line functions, network functions could also sit within a VNF, in which case a vSwitch may choose to classify the packet and switch it to this VNF. In order to provide service-to-service context and to preserve the initial classification of the packet, service chaining may be used to position more complicated functions inside or outside of a VNF.

The present document defines the critical aspects of vSwitch performance by treating the vSwitch as a Device Under Test (DUT), with specific configurations that are consistent across instantiations of a vSwitch on a compute platform. Existing testing and benchmarks specifications (see [i.1], [i.2], [i.3] and [i.4]) should be used to measure the performance of the DUT under these configurations and conditions, including measurement of metrics that support service engineering (such as the Composition Functions defined in IETF RFC 6049 [i.5]). The following configurations are of importance (see clause 7 for more detail and diagrams):

- vSwitch Physical to Physical: A vSwitch configured to receive traffic from a physical interface (uplink), make a forwarding decision, and re-forward the frame back out a physical interface (uplink).
- vSwitch Virtual to Virtual: A vSwitch configured to receive traffic from a VNF, make a forwarding decision, and re-forward the frame back out to a VNF.
- vSwitch VNF Loopback: A vSwitch is configured to receive traffic from a physical interface (uplink), make a forwarding decision, and then forward the frame to a VNF. The VNF should simply loopback the frames back to the vSwitch, which should do another forwarding decision to push the frame back out a physical interface (VNF).

In each configuration, the vSwitch may have a specific set of in-line functions configured such as L2 forwarding rules, L3 forwarding rules, tunnel termination, and wildcard rules used for ACLs, QoS, and monitoring. These in-line functions define the use case under test.

4.2 vSwitch Use Cases

4.2.1 Virtual Forwarding

The function of a vSwitch is minimally defined as a classification based on a locator (derived from the packet header) and a domain, both of which are matched upon to deliver the packet to a destination. The domain is derived either from the packet header (for example a VXLAN, or VNI) or from the ingress port on the vSwitch. The packet destination is either a port on the vSwitch, or a logical port that pushes the frames into a tunnel to send it to another vSwitch across the physical network. A vSwitch also needs to correctly handle broadcast, such that protocols such as ARP are correctly propagated between VNFs. Lastly, any packets not associated with current classification rules need to be handled in a specified 'default' manner, such as trapping them to the VIM or locally processing them to trigger additional classifications to be programmed into the vSwitch.

4.2.2 Overlay based Virtual Networks

Overlay networking separates the addresses in the underlay (physical) network from the addresses used by the VNFCs. A vSwitch does virtual forwarding to/from VNFCs on different vSwitches in the infrastructure over a tunnel connected point to point between the vSwitches.

Impacts of overlay protocols include:

- Overlay tunnels add bytes to each packet, which may cause the underlay network to fragment packets. Without adjusting the MTU used by the underlay and/or the VNFCs, fragmentation will occur, significantly impacting performance.
- The vSwitch will need to be able to look at the inner header of a tunnelled packet to do virtual forwarding, filtering, and in-line functions on the virtual network.
- Similarly, NICs in the physical compute node need to look at the inner headers as well to apply stateless offloads (RSS, checksums, TSO, etc.).

Since this feature can add to performance overheads, it is important to report the use of tunnels and their type when benchmarking. These tunnels may be a standard (for example VXLAN [i.6]), or they may be a custom or arbitrary tunnel depending on the environment. This may also include the use of encryption of the packets when frames are pushed into overlay tunnels.

4.2.3 Traffic Filtering

Traffic filtering can be abstracted into a match-action interface, where each filter matches on a set of conditions (such as fields on the frame, or internal metadata) which when matched will activate a set of actions. Traffic filters can be stateless or stateful. Stateful filtering may have performance challenges with respect to throughput, while stateless filtering might have performance challenges related to connection setup rate.

One use of filtering is a flow cache, where the complex traffic filtering operations are only applied to the first packet of a flow, and the result of the operations are recorded as an exact match flow cache entry. Subsequent packets from the same flow records a cache hit in the flow cache, eliminating the need for repeating the full filter set, and hence improving packets per second performance.

As traffic filtering in vSwitches represents an important set of use cases, the following performance metrics are key in evaluating vSwitch implementations:

- Connections established per second with various number of traffic filtering rules
- Packets per second with various number of concurrent connections
- Packets per second with various packet sizes and various number of traffic filtering rules
- Packets per second with different complexity of traffic filtering rules, including exact match and wild card rules
- Maximum concurrent connections
- CPU utilization without and with hardware based offloads for the above scenario

4.2.4 Distributed Network Services

Distributed network functions are built upon the same principles of filtering (stateless and stateful), and include:

- Distributed Routing
- Distributed firewall
- Distributed Virtual Load Balancer
- Network Address Translation (NAT)

Benchmarking of various vSwitch implementations needs to consider the above aspects. The following benchmarks are required with the following scenarios:

- Connection-establishment rate tests.
- Throughput tests with traffic across all flows with different packet sizes.
- Throughput tests with a high amount of control plane updates to the rules in the vSwitch.
- Combination test (throughput + Connection rate + Rule Updates).

4.2.5 Traffic Monitoring

Traffic monitoring as an in-line function of the vSwitch allows traffic to be monitored and sampled at the very edge of the network. In addition it may be valuable to sample this traffic using existing network monitoring formats and mechanisms.

Traffic mirroring capability mirrors traffic of virtual/logical/physical ports to special ports called "mirror ports" where traffic recorders/analysers are connected. Mirroring in the vSwitch can be expressed as an action activated by the filtering match, allowing mirroring to be active only on specific frames. Since traffic is classified and copied, performance of vSwitches depend on:

- The number of classification rules.

- The number of times traffic is copied.

Network monitoring may require sampling a subset of the matching traffic, as well as gather statistics at defined flow granularity levels and then export flow records to external collectors. Additional parameters that determine the effectiveness of monitoring include:

- Sampling frequency
- Number of observation points
- Number of simultaneous flows supported
- Number of records the monitoring switch can export per second
- Flow setup rate

4.2.6 Load Balancing

An individual VNF may have a maximum traffic load that it can handle, which may be traffic dependent and changing over time. In order to scale-out a VNF and abstract the granularity of the individual VNFs, traffic is distributed on a per-flow basis to multiple VNFC instances. The traffic distribution function needs to understand what constitutes a 'flow' or order preserving sequence of packets that carries packet-to-packet state within each VNF instance. This may vary depending on the type of VNF, however most VNFs will have similar definitions of flow (usually based on the N-tuple of the innermost IP and L4 fields):

- The distribution function is expected to be capable of receiving the aggregate traffic and distributing it. The aggregate traffic may be coming from a set of high capacity links (multiple 10G/40G/100G links).
- When a new VNF is added into the set of instances, the distribution of existing flows is not disturbed.
- When a VNF is removed from the set of instances, the distribution of existing flows is not disturbed.

4.2.7 Latency/Jitter Sensitive Workloads

Certain workloads require that the VNF process packets within a certain amount of time, and that the variation in this completion time is bounded in some way. In many cases the 'long tail' of the process completion time is the most important parameter for the workload, in that this worst case completion time may constitute a system-wide SLA failure. A latency and/or jitter (or delay variation) sensitive workload should be specified in one of the following ways:

- Average Completion Time: The average amount of time a VNF is allotted to complete its processing of a packet.
- Worst Case Completion Time: The worst case amount of time a VNF is allotted to complete its processing without being considered an error.
- Worst Case Completion Probability: The acceptable probability of a completion time outside of the upper bound.

For a given VNF, platform optimization and acceleration may be needed in order bring these completion times in-line with the requirements.

NOTE: The above addresses the VNF latency/jitter, a future revision will address the vSwitch components of latency/jitter and propose how to measure it.

For the purposes of the present document, the vSwitch performance for delay and delay variation needs to be specified and measured according to the metrics listed in clause 6.2 (for example, delay variation according to the PDV form in clauses 4.2 and 6.5 of IETF RFC 5481 [7], which can be adapted as a metric for VNF Process Completion Time as needed above).

4.2.8 Efficient Policy and QoS Control

Often as a result of platform optimization (for example, the use of SR-IOV) the ability for a policy control point to be in the path of the packets is lost, due to the increased latency/jitter, decreased bandwidth or increased processor load that this policy control point introduces. This gets to the core of the present document - the ability to benchmark the vSwitch applying this policy control and propose optimizations and/or accelerations to apply these policies without exceeding the fundamental performance requirements.

Policy control includes forwarding control as well as QoS (e.g. sharing data plane with packets of different policies) and performance requirements). QoS control usually involves applying different forwarding treatment to packets that are marked with pre-agreed code-points. The code-points correspond to classes of traffic, and the forwarding treatments can be applied on the basis of the code-point examination alone. Some classes of traffic may benefit from acceleration technologies, while for other classes acceleration may be unnecessary.

Therefore, the benchmarks listed in clause 6.2 need to be evaluated for systems implementing various policies and QoS forwarding treatments described above, including the ability to differentiate traffic and serve multiple policies at the same time.

4.2.9 Traffic Control & Traffic Shaping

It is always best if excessive incoming traffic is dropped as close to the network as possible.

vSwitches in each server are expected to take the job of controlling excessive traffic flowing in and out of VNFs. Since all the traffic going to the VNF is not of the same priority, some deployments require policing based on flow classification. Traffic flow classification can span across MAC, IP and transport headers and it could even be based on application protocol content.

Traffic Control as described above is for incoming traffic into the server either from the network or traffic that is looping between VNFs within the server. It is also expected that the traffic being sent out from VNF to other VNFs in the server or from VNFs to the network are shaped/conditioned so as to ensure that the local server does not overload other servers and network infrastructure. Also, shaping can be used to ensure that priority traffic is always accepted by upstream systems by sending higher priority traffic first.

Traffic Control and Traffic shaping with scheduling are expensive with respect to number of core cycles this functionality uses. Any vSwitch characterization and benchmarking needs to consider these two features. Some of the benchmarking metrics that are important to measure include:

- 5-tuple connection rate and packet rate with respect to number of virtual machines, virtual networks, scheduling algorithms, etc.

The benchmarks listed in clause 6.2 need to be evaluated for systems implementing Traffic Control and Traffic Shaping.

4.2.10 Flow Statistics Gathering

The vSwitch needs to be evaluated on statistics collection frequency, especially with a large number of flows.

4.2.11 Service Function Chaining

In Service Function Chaining (SFC) terminology, a VNF in the service chain is called a Service Function (SF). The vSwitches that participate in directing traffic to a set of SFs thus realizing the service chain are known as Service Function Forwarders (SFF). In the SFC use case, the vSwitch, as an SFF, plays a key role.

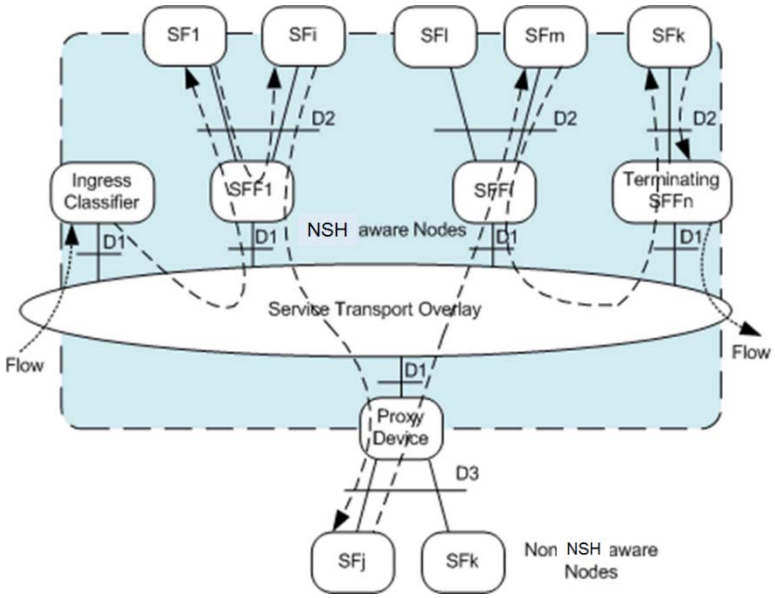

Figure 4.2.11-1: Service Function Chaining

In Figure 4.2.11-1 taken from the OPNFV VNFFG wiki each vSwitch/SFF forwards traffic to their local VNFs/SFs for service processing and then to the next hop SFF. The dotted lines represent the Network Forwarding Path or NFP which is an instance of a VNFFG or service function chain depicted in the figure. Since endpoint systems can be VMs, vSwitches are also required to do a classification function. An SFF may also terminate the chain by removing a service chain header and route the packet to its final destination.

Throughput and connection rate are key performance metrics of the vSwitch/SFF in SFC, especially as there can be multiple traversals between an SFF and SFs within the same compute node. Classification can also load share the connections across multiple SFs of same type. Furthermore, large number of simultaneous connections leads to very high number of flows in vSwitches. Hence, throughput with large number of flows also becomes an important metric.

Relevant metrics are listed below:

- Throughput and Packets per second with various packet sizes per service chain, in combination with the number of concurrent connections per service chain and overall across service chains.
- Throughput and Packets per second with different complexity of classification to service chain, including exact match and wild card rules.
- Maximum concurrent connections.
- Connection rate.

Concurrent with the metrics above, the following infrastructure resource readings are relevant:

- CPU utilization without and with hardware based offloads for the above scenario, along with the number of CPUs dedicated to the vSwitches/SFF and their configuration.
- Memory utilization with and without hardware based offloads.

5 Measurement Parameters

5.1 NFVI Host

To assure consistency in the execution and collection of performance benchmarks, the following characteristics of the underlying physical NFVI host need to be captured and documented.

NOTE: This clause draws on material from [i.4].

- Basic H/W platform details including model, configuration, revision
- CPU manufacturer, model, configuration (e.g. speed), stepping level, number of cores, hyper threading (or equivalent) status: enabled or disabled, cache size
- CPU microcode level
- BIOS version, release date, configuration options enabled
- Number of cores used for the test
- Memory information including type, size and speed
- Memory DIMM configuration, size, frequency and slot locations
- Number of physical NICs, and details including manufacturer, version, type, IO connectivity details (e.g. PCI slot installed)
- NIC Interrupt configuration
- Bus configuration parameters applicable (e.g. for PCI, payload size, early ACK options, etc.)
- Power management configuration at all levels (ACPI sleep states, processor package, O/S parameters applicable, etc.)

In parallel, the associated details of the NFVI software environment need to be captured (further drawing upon [i.4] for guidance):

- O/S and Kernel version
- O/S parameters and state of system under test (dedicated to test, shared with interactive users, etc.)
- Specified boot parameters, if any (e.g. provided to the Linux GRUB bootloader)
- Hypervisor details (type and version)
- Selected vSwitch, version number or commit id used
- vSwitch launch command line if it has been parameterized
- Memory allocation to the vSwitch; which NUMA node it is using; how many memory channels
- DPDK or any other SW dependency version number or commit id used
- Memory allocation to a VM - if it is from 'Huge pages' or elsewhere
- VM storage type: snapshot/independent persistent/independent non-persistent
- Number of VMs actively running
- Number of Virtual NICs (vNICs), versions, type and driver
- Number of virtual CPUs and their core affinity on the host
- Number vNIC interrupt configuration
- Thread affinity for the applications (including the vSwitch itself) on the host
- Details of Resource isolation, such as CPUs designated for Host/ Kernel (*isolcpu*) and CPUs designated for specific processes (*taskset*)

5.2 VNF

It is assumed that the process of on boarding and instantiation of a VNF will adequately capture the key parameters impacting the performance of a VNF, thus facilitating a consistent benchmarking.

Static elements such as the VNFD, and dynamic events stemming from VNF instantiation (e.g. the success or failure of resource allocation grants) are expected to capture the data required for consistent benchmarking.

6 Benchmarks

6.1 Environment

The environment of the system includes a traffic generator able to generate traffic consistent with the requirements in IETF RFC 2544 [2], a test receiver able to process packets and make the required measurements, plus the vSwitch under test. If a VNF loopback is to be used in the test, this loopback should also be instantiated in the platform. Since IETF RFC 2544 [2] procedures cause overload and extensive packet loss, they should only be use in an isolated lab environment (see IETF RFC 6815 [i.9]).

6.2 Traffic Profile

IETF RFC 2544 [2] throughput test shall be used to 'search' for the highest bandwidth that can be supported without loss. At the different offered load levels tested, the frame loss level is recorded. In scenarios with no packet loss, the latency and jitter of the system can be measured. However, IETF RFC 2544 [2] throughput test shall be augmented to produce metrics useful in performance engineering for services. The following is the list of recommended metrics:

- Mean One-way Delay (section 5.2 of IETF RFC 6703 [8])
- Minimum One-way Delay (sections 4 and 5.3 of IETF RFC 7679 [3])
- One-way Loss Ratio (sections 3 and 4.1 of IETF RFC 7680 [4])
- One-way Packet Delay Variation (according to the PDV form in sections 4.2 and 6.5 of IETF RFC 5481 [7])
- Packet Reordering Metrics (sections 3 and 4.1 IETF RFC 4737 [6])

IETF RFC 2544 [2] envisions tests over a range of packet sizes, and the requirements are no different here. Packets with small size (40 octets) will stress the packet transfer capabilities of computer systems, but higher bit rates can be achieved with large payloads. Packets of all sizes are present in the Internet. Following testing fixed sizes, it may be useful to conduct a few trials using a mix of sizes, commonly called an IMIX. The features of the IMIX should be specified according to IETF RFC 6985 [i.10].

Most benchmarking tests envision scenarios where multiple ingress ports (NICs) and egress ports are simultaneously utilized in order to assess the full capacity of the DUT. In practice, vSwitches will need to process traffic from both physical (P) and virtual (V) NICs simultaneously. Once the simple scenarios have been examined (P-P, P-V-V-P), other combinations should be examined in additional testing. The presence of shared resources shall be recognized and reported with test results, and scenarios should be designed where shared resources are minimized, or avoided completely.

The test streams generated in throughput testing should take on the characteristics of many individual flows. Each new flow may need to be processed and new policy established for forwarding subsequent packets. If a flow is idle for sufficient time, the process may need to be repeated.

When running with VNFs that go beyond simple loopback, it would be valuable to be able to test with stateful traffic (whose state is tracked by the acceleration technologies assisting the VNF itself and/or the vSwitch). This is considered outside the scope of this benchmark.

For deployment scenarios with IPsec tunnels, useful benchmarking and measurement methods are mentioned in [i.2], in particular see:

- Tunnel Capacity (section 8 of [i.2])

- Throughput of encryption and decryption operations (section 9 of [i.2])
- Latency of encryption and decryption operations (section 10 of [i.2])
- Tunnel setup rates (section 11 of [i.2])
- DoS attack Resiliency (section 15 of [i.2])

For deployment scenarios with Firewalls, the following specifications and their benchmarks shall be measured and reported. These benchmarks are relevant owing to their reliance on the vSwitch as well as the Firewall VNF under test:

- IP Throughput and Maximum Forwarding Rate, section 5.1 of IETF RFC 3511 [5]
- DoS on TCP Connection Establishment rate and HTTP transfer rates, section 5.5 of IETF RFC 3511 [5]
- Maximum HTTP transaction rate, section 5.7 of IETF RFC 3511 [5]
- IP Fragmentation handling, section 5.9 of IETF RFC 3511 [5]
- Latency, section 5.10 of IETF RFC 3511 [5]

For the VBS use case, with a focus on comparing the connectivity of either attached or independent implementations of acceleration technology, the benchmarks listed above are applicable (with the necessary attention to configuration). To fully assess the VBS system, and be able to compare virtual implementations based on attached and independent architectures, and possibly extend the comparison to physical BS implementations, benchmarks are needed for the BS System Under Test (SUT). The ability to easily compare physical and virtual implementations is a key consideration listed in [i.3].

For systems implementing Policy and QoS Control, the terms and benchmarks of ETSI GS NFV-INF 007 [i.7] are relevant, in addition to the list of metrics intended to augment IETF RFC 2544 [2] above.

For systems implementing Traffic Control and Traffic Shaping, the benchmarks described in clause 6 of ETSI GS NFV-IFA 002 [i.8] are relevant.

7 Deployment Scenarios

7.1 Use Case Example

A use case can be classified by the following parameters:

Parameter Category	Parameters and Values for the Scenario
Attach	Physical Network Attach: _____ Virtual Network Attach: _____
Infrastructure	The configuration of the compute node such as number of cores used for the vSwitch, the core frequency, software versions and physical infrastructure specs and resources (NIC, storage, memory, etc.). See clause 5.
Switching	Tunnel Type(s):_____ Locator(s): _____ Domain(s): _____
Filtering	Stateless Filter Matches:_____ Stateful Filter Matches: _____ Filter Actions: _____
In-line Functions	List of additional in-line functions, and how they are configured when the system was benchmarked. Examples: Flow Mirroring/Sampling, QoS traffic shaping, in-line firewall.
Chained Functions	List of additional chained functions that are part of the infrastructure before packets transit to/from the physical and virtual networks (see Datapath Figures in clause 7.2).
Attached VNFs	At a minimum, some number of VNFCs need to be instantiated to exercise the virtual network attach ports. A vloop VNFC receives traffic and re-transmits it back out either unchanged or its destination MAC changed to a given NextHop (see Datapath Figures in clause 7.2).

Parameter Category	Parameters and Values for the Scenario
Test Stimuli	IETF RFC 2544 Offered Loads { } Packet Size Distributions { } Connections per second during Benchmark: _____ Max concurrent: _____ New Connection Rate _____ Sustained vSwitch table entries during benchmark _____ Number of statistics gathered / sec during Benchmark: _____
Measured Results	Physical Network (P-P) Packets/sec _____Latency: Max: _____ Avg: _____ Min: _____ Virtual Network (P-V-V-P) Packets/sec: _____ Latency: Max: _____ Avg: _____ Min: _____ CPU Utilization During Benchmark: _____

7.2 Virtual Switch Datapath

The virtual switch can be configured and assessed in its most elemental form using the physical interface to physical interface (P) loop through the vSwitch (vSw) as shown in Figure 7.2-1.

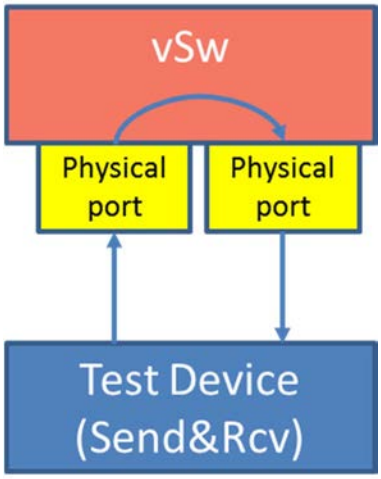

Figure 7.2-1: P - vSw - P

Note that two physical interfaces may be used, or connectivity may be provided by a single physical interface in full duplex mode.

Because virtual interfaces have different performance implications, a test set-up including a simple VNF can be used to determine the performance and latency associated with a path through the virtual interfaces. The datapath in Figure 7.2-2 would encounter two virtual interfaces and a trivial VNF set to loop the packet/frame through to the next virtual interface.

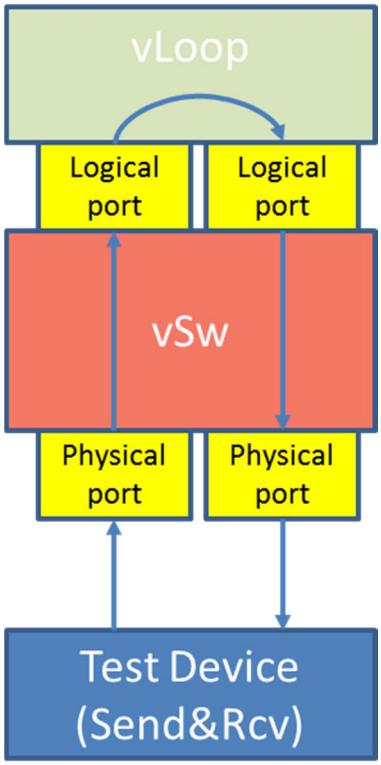

Figure 7.2-2: P - vSw - (vLoop VNF) - vSw - P

The vLoop is intended to be the same image, whether testing with or without acceleration. Also the vLoop needs to be implanted so that it is not the bottleneck limiting Throughput testing.

With an overlay network included, the datapath is as shown in Figure 7.2-3.

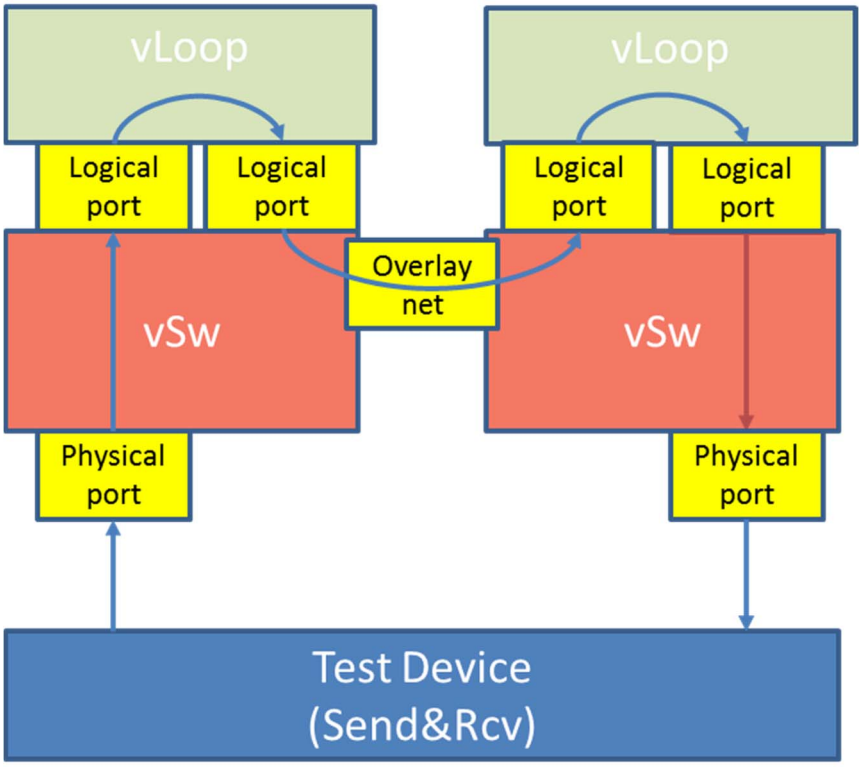

Figure 7.2-3: P - vSw - (vLoop VNF) - vSw - Overlay Network - vSw - (vLoop VNF) - vSw - P

When acceleration technologies are applied to these simple configurations, the path needs to identify how the path changes to include acceleration, as discussed in clause 7.3.

7.3 Acceleration Datapath

This clause considers the modifications to the datapath when incorporating acceleration and in order to make fair comparisons with the non-accelerated datapaths described in clause 7.2.

When using acceleration in comparison with a non-accelerated path, there should be minimal modifications to attempt to ensure a fair comparison between accelerated and non-accelerated configurations.

In general, vSwitch acceleration can be included as shown in Figure 7.3-1, where the vSwitch operation itself is accelerated through one of a number of means.

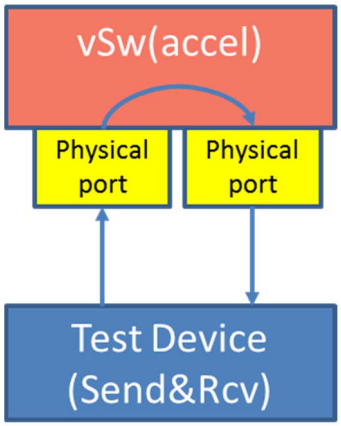

Figure 7.3-1: P - vSw(accel) - P

Figure 7.3-2 indicates individual placements of acceleration technology.

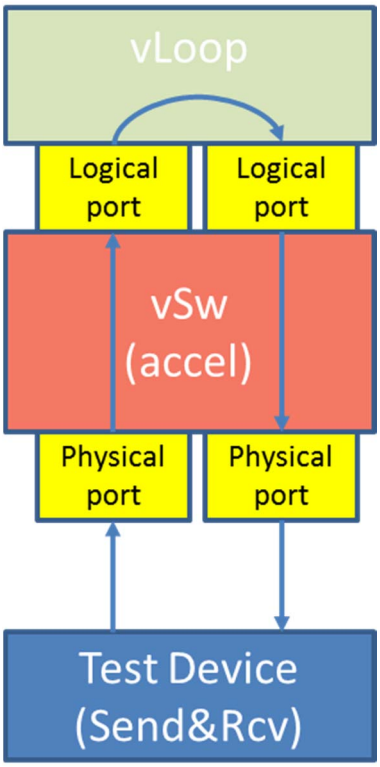

Figure 7.3-2: P - vSw(accel) - (vLoop VNF) - vSw(accel) - P

In Figure 7.3-3, one host does not have acceleration technology:

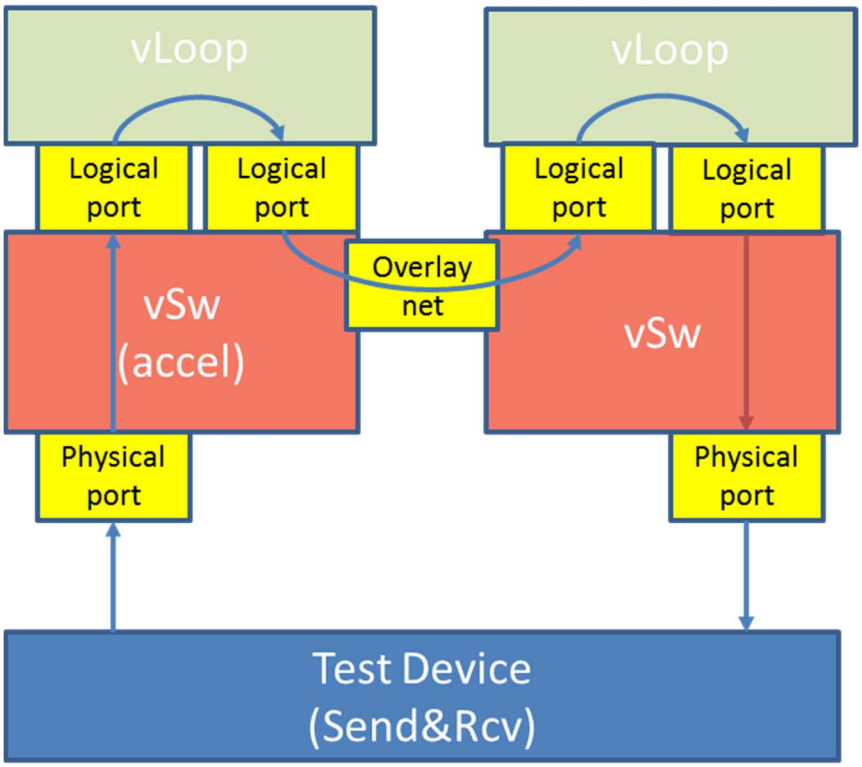

Figure 7.3-3: P - vSw(accel) - (vLoop) - vSw(accel) - Overlay Net - vSw - (vLoop) - vSw - P

8 Follow-on PoC Proposals

The OPNFV vSwitch Performance Project (VSPERF) [i.4] has begun to implement test specifications that meet many of the requirements in the present document, and have considered additional test specifications for development based on the remaining requirements specified here. Further, the VSPERF project aims to deliver an automated testing platform as part of OPNFV releases, implementing their test specifications and therefore, the requirements of the present document. Therefore, the early adoption and implementation of the present document is considered more valuable than a simple PoC, in that an Open Source tool and test specifications are lasting products that the NFV industry can use.

Annex A (informative): Authors & contributors

The following people have contributed to the present document:

Rapporteur:
Rabi Abdel Hafiz, Vodafone Group

Other contributors:
Nabil Damouny, Netronome

Al Morton, AT&T Labs

Srini Addepalli, Intel

Ron Breault, Wind River

Dan Daly, Intel

François-Frédéric Ozog, 6WIND

Annex B (informative): Bibliography

ETSI GS NFV-INF 003: "Network Functions Virtualisation (NFV); Infrastructure; Compute Domain".

ETSI GS NFV-INF 004: "Network Functions Virtualisation (NFV); Infrastructure; Hypervisor Domain".

ETSI GS NFV-SWA 001: "Network Functions Virtualisation (NFV); Virtual Network Functions Architecture".

History

Document history		
V2.1.1	April 2016	Publication
V2.3.1	August 2017	Publication