ETSI GS NFV-TST 009 V3.2.1 (2019-06)

Network Functions Virtualisation (NFV) Release 3; Testing; Specification of Networking Benchmarks and Measurement Methods for NFVI

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-TST009ed321

Keywords
benchmarking, measurement, NFV, NFVI

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

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Copyright Notification

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables 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.

Introduction

The widespread adoption of virtualised implementation of functions has brought about many changes and challenges for the testing and benchmarking industries. The subjects of the tests perform their functions within a virtualisation system for additional convenience and flexibility, but virtualised implementations also bring challenges to measure their performance in a reliable and repeatable way, now that the natural boundaries and dedicated connectivity of physical network functions are gone. Even the hardware testing systems have virtualised counterparts, presenting additional factors to consider in the pursuit of accurate results.

The present document draws on learnings from many early benchmarking campaigns and years of benchmarking physical network functions to develop and specify new normative benchmarks and methods of measurement to characterize the performance of networks in the Network Function Virtualisation Infrastructure.

1 Scope

The present document specifies vendor-agnostic definitions of performance metrics and the associated methods of measurement for Benchmarking networks supported in the NFVI. The Benchmarks and Methods will take into account the communication-affecting aspects of the compute/networking/virtualisation environment (such as the transient interrupts that block other processes or the ability to dedicate variable amounts of resources to communication processes). These Benchmarks are intended to serve as a basis for fair comparison of different implementations of NFVI, (composed of various hardware and software components) according to each individual Benchmark and networking configuration evaluated. Note that a Virtual Infrastructure Manager (VIM) may play a supporting role in configuring the network under test. Examples of existing Benchmarks include IETF RFC 2544 [1] Throughput and Latency (developed for physical network functions).

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 https://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] IETF RFC 2544 (March 1999): "Benchmarking Methodology for Network Interconnect Devices".

[2] IETF RFC 2285 (February 1998): "Benchmarking Terminology for LAN Switching Devices".

[3] IETF RFC 2889 (August 2000): "Benchmarking Methodology for LAN Switching Devices".

[4] IETF RFC 6985 (July 2013): "IMIX Genome: Specification of Variable Packet Sizes for Additional Testing".

[5] ETSI GS NFV-TST 008 (V3.1.1): "Network Functions Virtualisation (NFV) Release 3; Testing; NFVI Compute and Network Metrics Specification".

[6] ETSI GS NFV 003 (V1.3.1): "Network Functions Virtualisation (NFV); Terminology for Main Concepts in NFV".

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] ETSI GS NFV-INF 003 (V.1.1.1): "Network Functions Virtualisation (NFV); Infrastructure; Compute Domain".

[i.2] Bregni, Stefano, (October 1996): "Measurement of Maximum Time Interval Error for Telecommunications Clock Stability Characterization", IEEE Transactions on Instrumentation and Measurement.

NOTE: Available at http://home.deib.polimi.it/bregni/papers/mtiemeas.pdf.

[i.3] ETSI GS NFV-TST 001 (V1.1.1): "Network Functions Virtualisation (NFV); Pre-deployment Testing; Report on Validation of NFV Environments and Services".

[i.4] FD.io VPP Developer Documentation: "Shared Memory Packet Interface (memif) Library".

NOTE: Available at https://docs.fd.io/vpp/17.10/libmemif_doc.html.

[i.5] FD.io VPP pma-tools: "Software tools for performance and efficiency measurements".

NOTE: Available at https://git.fd.io/pma_tools/tree/jitter.

[i.6] John D. McCalpin: "STREAM: Sustainable Memory Bandwidth in High Performance Computers".

NOTE: Available at https://www.cs.virginia.edu/stream/.

[i.7] Intel Software Developer Zone: "Intel® Memory Latency Checker v3.5".

NOTE: Available at https://software.intel.com/en-us/articles/intelr-memory-latency-checker.

[i.8] IETF RFC 8172 (July 2017): "Considerations for Benchmarking Virtual Network Functions and Their Infrastructure".

[i.9] IETF RFC 8204 (July 2017): "Benchmarking Virtual Switches in the Open Platform for NFV (OPNFV)".

[i.10] Cooper, T., Morton, A., Rao, S., OPNFV Summit Presentation: "Dataplane Performance, Capacity, and Benchmarking in OPNFV", June 2017.

NOTE: Available at https://wiki.opnfv.org/download/attachments/10293193/VSPERF-Dataplane-Perf-Cap-Bench.pptx?api=v2.

[i.11] IETF RFC 5481 (March 2009): "Packet Delay Variation Applicability Statement".

[i.12] Wikipedia: "Binary Search".

NOTE: Available at https://en.wikipedia.org/wiki/Binary_search_algorithm.

[i.13] GeeksforGeeks Computer Science Portal: "Searching Algorithms, Binary Search".

NOTE: Available at https://www.geeksforgeeks.org/binary-search/ (including code implementations).

[i.14] Andrzej Pelc: "Searching games with errors - fifty years of coping with liars", Theoretical Computer Science 270 (2002) pp 71-109.

NOTE: Available at https://www.gwern.net/docs/statistics/comparison/2002-pelc.pdf.

[i.15] Presentation to IETF-102 BMWG Session: "Evolution of Repeatability in Benchmarking: Fraser Plugfest (Summary for IETF BMWG)", Sridhar Rao, Al Morton, and OPNFV VSPERF Project Team.

NOTE: Available at https://datatracker.ietf.org/meeting/102/materials/slides-102-bmwg-evolution-of-repeatability-in-benchmarking-fraser-plugfest-summary-for-ietf-bmwg-00.

[i.16] LightReading/EANTC NFV Tests and Trials: "Validating Cisco's NFV Infrastructure", October 2015.

NOTE: Available at http://www.lightreading.com/nfv/nfv-tests-and-trials/validating-ciscos-nfv-infrastructure-pt-1/d/d-id/718684.

[i.17] OPNFV NFVbench Project Description.

NOTE: Available at https://wiki.opnfv.org/display/nfvbench.

[i.18] Python Package Index MLResearch 0.1.1.0: "Library for speeding up binary search using shorter measurements".

NOTE: Available at https://pypi.org/project/MLRsearch/ and https://docs.fd.io/csit/master/report/introduction/methodology.html#mlrsearch-algorithm.

[i.19] Recommendation ITU-T Y.1541 (December 2011): "Network performance objectives for IP-based services".

[i.20] IETF RFC 8337 (March 2018): "Model-based Metrics".

[i.21] IETF RFC 7348: "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualised Layer 2 Networks over Layer 3 Networks".

[i.22] OPNFV VSPERF Long Duration Testing.

NOTE: Available at https://wiki.opnfv.org/display/vsperf/Long+Duration+Testing.

[i.23] OPNFV VSPERF Cross-NUMA Node Testing.

NOTE: Available at https://wiki.opnfv.org/display/vsperf/Cross-NUMA+performance+measurements+with+VSPERF.

3 Definition of terms, symbols and abbreviations

3.1 Terms

For the purposes of the present document, the terms given in ETSI GS NFV 003 [6] and the following apply:

NOTE: A term defined in the present document takes precedence over the definition of the same term, if any, in ETSI GS NFV 003 [6].

burst: stream of two or more frames transmitted with the minimum allowable inter-frame gap, as in a burst of frames

bursty traffic rate: stream consisting of repeated bursts that maintains a specified frequency of transmitted frames per second, such that the frequency equals the reciprocal of the sum of the constant inter-burst gap and the burst serialization time (for a constant frame size and minimum allowable inter-frame gaps between frames in the burst)

NOTE: See section 21 of IETF RFC 2544 [1].

constant frame rate stream: stream that maintains a specified frequency of transmitted frames per second, such that the frequency equals the reciprocal of the sum of the constant inter-frame gap and the frame serialization time (for a constant frame size)

flow: set of frames or packets with the same n-tuple of designated header fields that (when held constant) result in identical treatment in a multi-path decision (such as the decision taken in load balancing)

frame size: fixed length of a frame in octets (or 8-bit bytes), all headers included

NOTE: For example, Ethernet frame size includes the frame CRC, but exclude the transmission overhead per frame of 20 octets (the preamble and the inter frame gap).

measurement goal: specific criteria that a measurement result is expected to meet to satisfy the requirements of a benchmark definition

method: series of one or more Sets of Tests conducted to achieve a measurement goal

offered load: both the count (in frames) and transmission rate (in frames per second) generated by the measurement system during a trial, including both directions of transmission with bi-directional streams

pod: partition of a compute node that provides an isolated virtualised computation environment, for one or more virtualisation containers in the context of an Operating System Container virtualisation layer

set: series of one or more tests conducted to achieve a measurement goal

stream: population of frames or packets with various header attributes, that contain one or more flows and comprise the Offered Load

termination criterion: for Long Duration Testing, the set of Trial outcomes which constitute the conditions for early termination of a Test (because the impairments are excessive and the intended purpose of testing will likely not be met)

test: series of one or more trials conducted to achieve a measurement goal

trial: single iteration of a measurement producing one or more results that can be compared with search termination criteria

3.2 Symbols

Void.

3.3 Abbreviations

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

ARP	Address Resolution Protocol
AV	Array with loss Verification
BIOS	Basic Input Output System
BLV	Binasry search with Loss Verification
BSwLV	Binary Search with Loss Verification
CPU	Central Processing Unit
CRC	Cyclic Redundancy Check
DCI	Data Center Interconnect
DUT	Device Under Test
DV	Delay Variation
FDV	Frame Delay Variation
GRE	Generic Routing Encapsulation
IFDV	Inter-Frame Delay Variation
IMIX	Internet Mix (of frame or packet sizes)
IP	Internet Protocol
L2	Layer 2
MTIE	Maximum Time Interval Error
NA	Not Available
NDR	No Drop Rate
NFV	Network Function Virtualisation
NFVI	Network Function Virtualisation Infrastructure
NIC	Network Interface Card
NSH	Network Service Header
NUMA	Non-Uniform Memory Access
OPNFV	Open Platform for NFV
OS	Operating System
OVS	Open VSwitch
PCIe	Peripheral Component Interconnect express
PDR	Partial Drop Rate
PMD	Poll Mode Driver
PVP	Physical Virtual Physical (Single VNF)
PVVP	Physical Virtual Virtual Physical (Multiple VNF)
SDN	Software Defined Network
SFC	Service Function Chains
SR-IOV	Single Root - Input Output Virtualisation
SRv6	Segment Routing over IPv6 dataplane
SUT	System Under Test

TCP	Transmission Control Protocol
Thpt	Throughput
VethX	Virtual Ethernet Interface
VIM	Virtual Infrastructure Manager
VLAN	Virtual Local Area Network
VM	Virtual Machine
VNF	Virtual Network Function
VNFC	VNF Component
vSw	virtual Switch
VSPERF	vSwitch Performance Project in OPNFV
VTEP	Virtual extensible local area network - Tunnel End Point
VXLAN	Virtual eXtensible Local Area Network

NOTE: See IETF RFC 7348 [i.21].

VXLAN-GPE	Virtual eXtensible Local Area Network - Generic Protocol Extension

4 Time and Time Intervals for Metrics and Benchmarks

In the present document, coherent compute domains comprise the System Under Test (SUT) [i.3] which contains one or more Devices Under Test (DUT), the platform for software-based measurement systems, and may also provide the automation platform for overall SUT, DUT, and test measurement system installation and configuration. The requirements of this clause are particularly relevant to test devices or test functions (such as the Traffic Generator or Receiver), and are generally referred to as measurement systems.

Coherent compute domains [i.1] usually need access to a clock with accurate time-of-day (or simply date-time) and sources of periodic interrupts. Time sources are accessed to provide timestamps for events and log entries that document the recent history of the compute environment. Periodic interrupts provide a trigger to increment counters and read current conditions in the compute and networking environments. The compute domain may contain a very large number of NFV compute nodes [i.1], and each node needs to execute a process to synchronize its hardware and system clocks to a source of accurate time-of -day, preferably traceable to an international time standard.

With the foundation of time, date, and periodic interrupts, a measurement system can determine the beginning and end of time intervals, which is a fundamental aspect of metrics that involve counting and collecting events (such as frame or packet transmission and reception), and a SUT or DUT can provide accurate event logging and other functions.

Table 4-1 specifies requirements applicable to time, date, and periodic interrupts for all systems, and includes an additional Requirement for Test Devices/Functions (General-Time-03).

Table 4-1: Requirements applicable to time, date and periodic interrupts

General-Time-01	Each node in the compute domain shall be able to take readings from (or access) a clock with accurate time-of-day and calendar date, having less than ± one millisecond error from an international time standard over a 900 second measurement interval.
General-Time-02	Each node in the compute domain shall have a source of periodic interrupts available which are derived from the time-of-day clock, with configurable period (a parameter of metrics that use this feature).
General-Time-03	The Maximum Time Interval Error (MTIE) between the local clock of the Test Device/Function and a reference time source shall be specified in terms of S, the observation interval, and TE, the maximum peak-to-peak deviation of time error. S is provisionally set at 120 seconds [i.2].

5 Framework for Metric and Benchmark Definitions

The metric and Benchmark definitions in the present document are primarily derived from the industry precedents established when IETF RFC 2544 [1] was first published.

For each metric and Benchmark it specifies, the present document provides the following template of elements in separate sub-clauses, many of which are required for reporting:

- Background
- Name
- Parameters (input factors)
- Scope of coverage (possibly a subset of resources tested)
- Unit(s) of measure
- Definition
- Method of Measurement (unique aspects, in addition to clause 11)
- Sources of Error
- Discussion
- Reporting Format

NOTE: The present document specifies Benchmarks and metrics, some of which are well-known by name, but whose definitions and other details are presented as normative requirements in the present document, according to the framework of elements above. See clauses 8 and above for the current implementation of this framework and template elements.

Where a clause specifies both a Benchmark and one or more metric-variants, the tester shall measure and report the Benchmark along with optional metric-variants in all testing campaigns claiming compliance with the present document.

6 Test Set-ups and Configuration

6.1 Goals of Benchmarking and Use Cases

The use cases supported by the testing effort shall be clearly identified, in order to select the applicable set of test setups. For example (referring to setups in clause 6.2, figure 6.2-1), testing the firewall-related functions of a vSwitch (using match-action rules) can be accomplished using the Phy2Phy setup, while a virtualised web host would typically require the PVP setup, and a chain of payload-processing VNFs (transcoding and encryption) would require the PVVP setup. The following list contains important considerations for Test Setups:

1) The use case(s) addressed by the goals of the test will help to determine whether tests with streams composed of multiple flows are necessary, and how the streams should be constructed in terms of the header fields that vary (such as address fields).

2) The use case(s) addressed by the goals of the test will help to determine what protocol combinations should be tested, and how the streams should be constructed in terms of the packet header composition and mix of packet sizes [4]. There may be one or more unique frame size associated with the particular use case of interest, and these may be revealed using traffic monitoring as described in IETF RFC 6985 [4].

3) At least two mixes of packet size should be specified for testing, sometimes called "IMIX" for Internet mix of sizes. IETF RFC 6985 [4] provides several ways to encode and specify mixes of packet sizes, and the RFC encourages repeatable packet size sequences by providing specifications to support reporting and reproducing a given sequence (called the IMIX Genome).

4) Bidirectional traffic should possess complimentary stream characteristics (some use cases involve different mixes of sizes in each direction).

5) Protocols like VLAN, VXLAN, GRE, VXLAN-GPE, SRv6 and SFC NSH are needed in NFVI deployments and some will require multiple encapsulations (VLAN and VXLAN).

6.2 Test Setups

The following general topologies for test setups should be used when called for by the target use cases. The test designers shall prepare a clear diagram of their test setup, including the designated addresses (Layer 2 and/or IP subnet assignments, if used). This diagram will help trouble resolution in a setup with connectivity issues.

Figure 6.2-1 illustrates the connectivity required for test device/functions, NFVI components, and testing-specific VNFs. Note that arrows indicate one direction of transmission (one-way), but bi-directional testing shall be performed as required. Also, the data plane switching/forwarding/routing function is illustrated in red and labelled "vSw". The number of Physical port pairs in use may also be expanded as required (only one pair is shown).

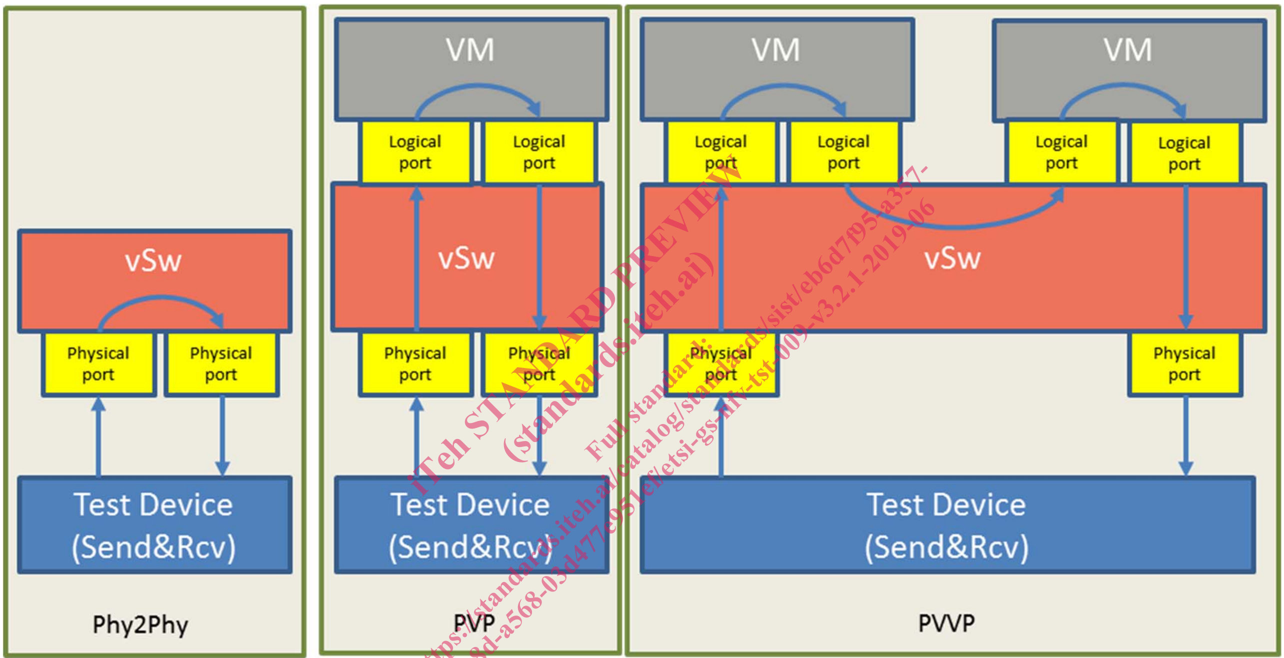

Figure 6.2-1: General Dataplane Test Setups in a single host: Baseline(Phy2Phy), Single VNF (PVP), Multiple VNF (PVVP)

A control-protocol may be used between the "vSw" and an SDN controller, but this control plane connectivity is not illustrated in figure 6.2-1.

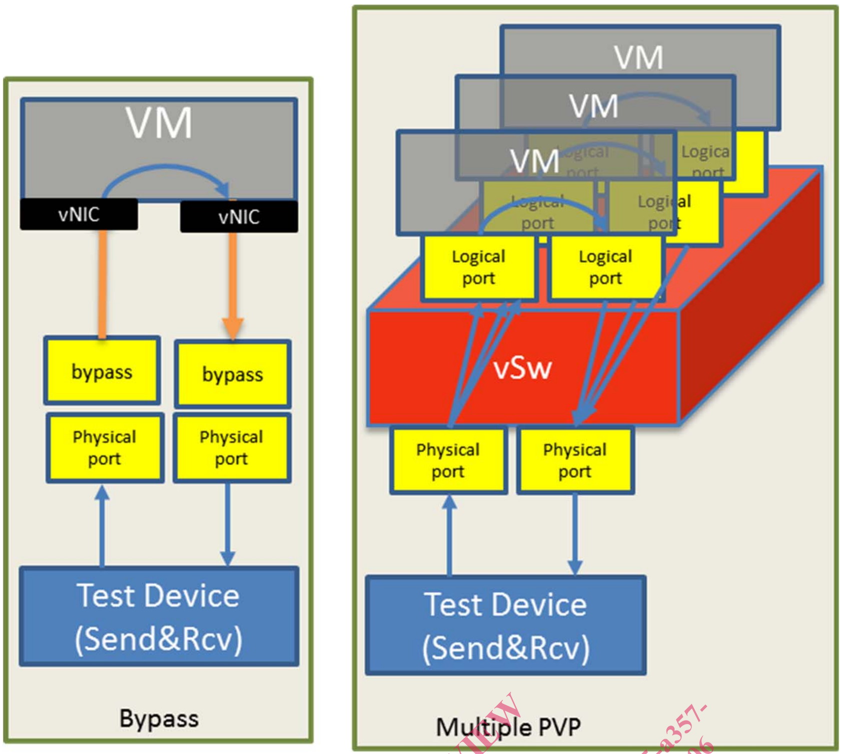

Figure 6.2-2: Additional Data plane Test Setups in a single host: Bypass and Multiple PVP

Figure 6.2-2 illustrates additional test setups, where Bypass permits testing of technologies such as Single-Root Input Output Virtualisation (SR-IOV), and Multiple PVP accommodates testing with many parallel paths through VMs. Of course, the Multiple VNF setup (PVVP) can be expanded with additional VMs in series. Mesh connectivity could also be applied between the vSwitch and multiple physical port interconnections, similar to the one to many connectivity illustrated in the Multiple PVP scenario in figure 6.2-2. In the limit, many test setups could be deployed simultaneously in order to benchmark a "full" SUT, with or without CPU oversubscription.

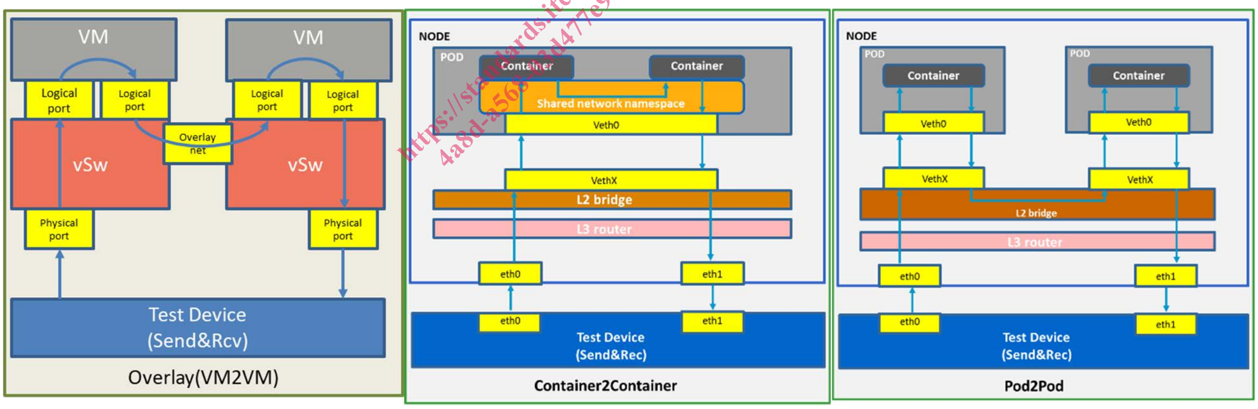

Figure 6.2-3: Additional Data plane Test Setups in a single host: Overlay(VM2VM), Container2Container and Pod2Pod

Figure 6.2-3 illustrates some of the unusual connectivity scenarios that are possible, using overlay networks to connect between VMs, the capability of namespace networking for Container to Container communications in Operating Systems Containers, or the L2 bridge provided by Container Orchestration System for Pod to Pod communication in the same node. Bypass technologies for more direct access are also possible with Containers (SR-IOV), and Container to Container communication may take place through shared packet memory [i.4].

In a common Container Infrastructure, the node represents a physical host. The pod represents the basic unit for orchestration and management which is able to host multiple containers. For internal communication, the L2 bridge is used to connect all the pods. For external communication, the L3 router may be used to manage the traffic in and out of the node.