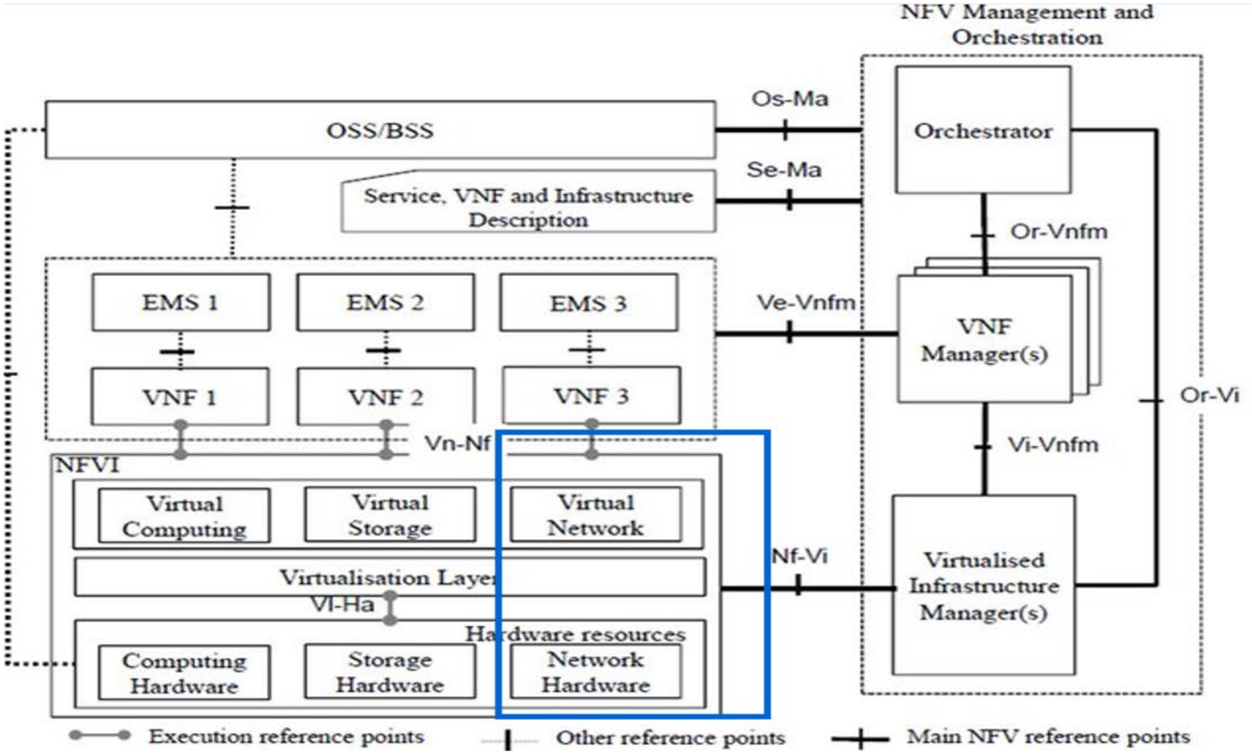
Identified networking gaps

When looking into the existing ETSI NFV Architectural Framework[1)] model as depicted in the figure below there are a number of gaps identified to enable a useful Reference Model for the Networking

1. ETSI NFV does not have a separation in between HW Infrastructure Management and SW Virtualization Management and by that the cardinality of having multiple CaaS and IaaS layers on top of a shared HW Infrastructure Layer cannot be expressed
2. ETSI NFV lacks a description of the reference points in between the SW Virtualization Layers and the HW Infrastructure Layer denoted as Vl-Ha and by that cannot express Packet Flows, Control/Status Interfaces and Management Interfaces them in between
3. ETSI NFV model does not explicitly call out an HW Abstraction layer on top of the HW resources and by that many SW Virtualization Layers act upon the physical HW components in a non-portable way that do not allow an HW Infrastructure to compose abstracted HW resources from pools of individual HW components
4. ETSI NFV lack of spelled out HW Abstraction does not enable a clear way to abstract and describe the technical possibilities that HW layer networking separation, encapsulation, acceleration, etc. can be implemented either in a HW Infrastructure Managed SmartNIC on the Server units or in the shared underlay switching fabric i.e. on the physical Switching Units
5. ETSI NFV model does not include any reference to SDN (Software Define Network) controller(s) and relevant integration points into NFVI and MANO
6. ETSI NFV model does not have a way to enable programmable forwarding planes in the HW layer controlled from higher layers of virtualization managers, orchestrators or Network Functions



Layering and Concepts

Cloud and Telco networking are layered, and it is very important to keep the layering dependences low to enable security, separation and portability in between multiple implementations.

These explanations for Layering and Concepts are likely too rich and verbose for a CNTT document entry, but it is important that we first understand and agree on these concepts before we start entering text that could be misunderstood.

Underlay and Overlay Networking concepts

The ETSI NFV model divide networking in an Underlay and an Overlay Network layer. The purpose with this layering is to ensure separation of the SW Virtualization tenants Overlay Networks from each other, whilst allowing the traffic to flow on the shared Underlay Network in between all Ethernet connected HW units.

The Overlay Networking separation is often done through encapsulation e.g. through VxLAN on the Underlay Networks e.g. based on L2 (VLAN) or L3 (IP) networks.

In some instances, the SW Virtualization Tenants can bypass the Overlay Networking encapsulation to achieve better performance or network visibility/control. A common method to bypass the Overlay Networking encapsulation is the usage of SR-IOV that effectively hands up the NIC Physical and Virtual Functions on the NIC to the SW Virtualization Layer and Tenants. In these cases, the Underlay Networking must handle the separation e.g. through a Virtual Termination End Point (VTEP) that encapsulate the Overlay Network traffic.

Software Defined Networking control concepts of the Underlay Networking

VTEP could be manually provisioned in the Underlay Networking or be automated and controlled through a Software Defined Networking interfaces to the Underlay Networking in the HW Infrastructure Layer. Due to the many different facets of Software Defined Networking we will here denote them SDN Underlay (SDNu).

When there are multiple simultaneous SW Virtualization Layers on the same HW Infrastructure, there is a need to ensure Underlay networking separation in the HW Infrastructure Layer. This separation can be done manually through provisioning of a statically configured separation of the Underlay networking in the HW Infrastructure Layer. A better and more agile usage of the HW Infrastructure is to have an authoritative SDN provisioning controller function (here denoted SDNuP) that can be controlled through an automation interface from a HW Infrastructure Orchestrator. The main tasks for the SDNuP are to discover and establish the Underlay resources and the ensure separation of shared HW Infrastructure Underlay networking resources.

Multiple Containerized Virtualization Layer (CaaS) running on an Infrastructure as a Service (IaaS) Virtualization Layer could make use of the IaaS layer to handle Underlay Networking separation. In these cases, also the IaaS Virtualization Infrastructure Manager (VIM) could include a SDNu control interface enabling automation.

Hardware and Software Infrastructure Layer concepts

For Cloud implementations of multiple well separated simultaneous SW Virtualization domains on a shared HW Infrastructure there must be a separation of the hardware resources e.g. servers and the Underlay Networking resources that interconnect the hardware resources e.g. through a switching fabric.

To allow multiple separated simultaneous SW Virtualization domains onto a shared switching fabric there is a need to split up the Underlay Networking resources into non overlapping addressing domains on suitable protocols e.g. VxLAN with their VNI Ranges. This separation must be done through an administrative domain that could not be compromised by any of the individual SW Virtualization domains either by malicious or unintentional Underlay Network separation configuration.

The ETSI NFVI Infrastructure Management thus must be split into two distinct layers that can be referred to as HW Infrastructure Layer and SW Infrastructure Layer which can be managed separately from different administrative domains. When there are multiple separated simultaneous SW Virtualization domains, they have to be possible to be individual administrative domains.

1) Referenced ETSI NFV model in the Architectural Framework: https://www.etsi.org/deliver/etsi_gs/NFV/001_099/002/01.02.01_60/gs_NFV002v010201p.pdf

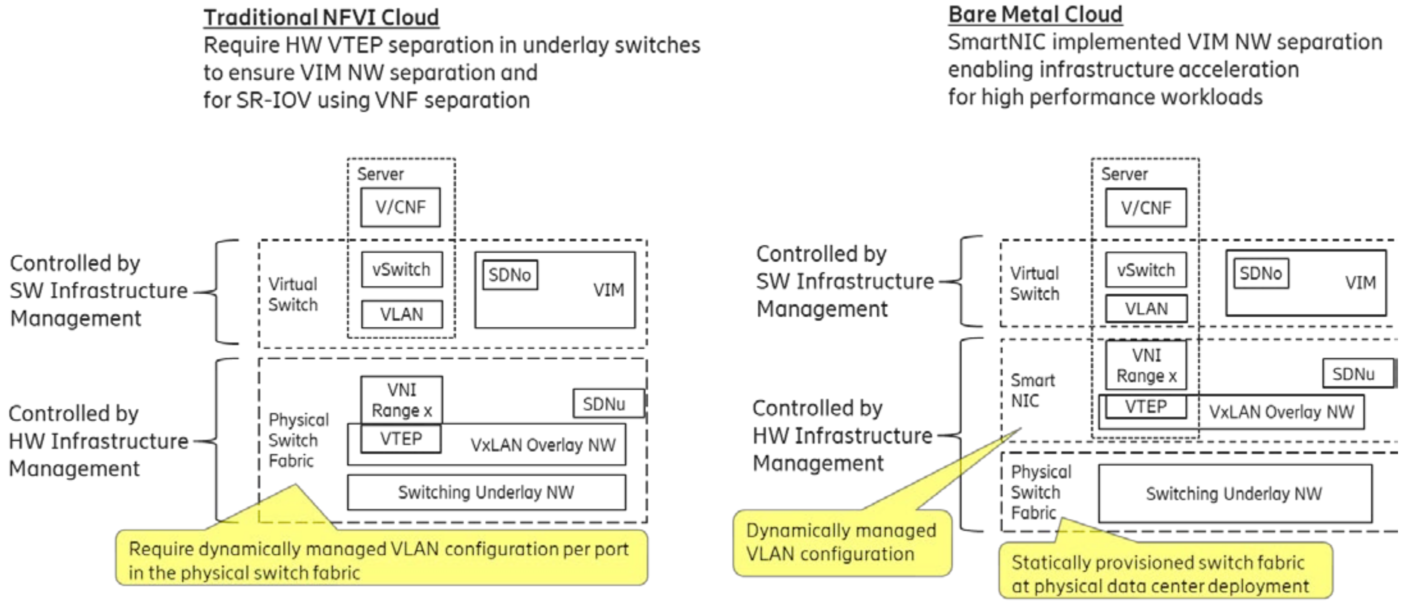
Switch Fabric and SmartNIC concepts for Underlay Networking separation

The HW Infrastructure Layer can implement the Underlay Networking separation in any type of packet handling component. This may be deployed in many different ways depending on target use case requirements, workload characteristics and available platforms. Two of the most common ways is 1.

within the physical Switch Fabric and 2. in a SmartNIC connected to the Server CPU being controlled over a management channel that is not reachable from the Server CPU and its host software. In either way the Underlay Networking separation is controlled by the HW Infrastructure Manager.

In both cases the Underlay Networking can be externally controlled over the SDNu interface, that must be instantiated with appropriate Underlay Networking separation for each of the SW Virtualization administrative domains.

Two exemplifications of different common HW realizations of Underlay separation in the HW Infrastructure Layer can be seen in the figure below.


Traditional NFVI Cloud
Require HW VTEP separation in underlay switches to ensure VIM NW separation and for SR-IOV using VNF separation
Server
V/CNF
Controlled by SW Infrastructure Management
Virtual Switch
vSwitch
VLAN
SDNo
VIM
Controlled by HW Infrastructure Management
Physical Switch Fabric
VNI Range x
VTEP
SDNu
VxLAN Overlay NW
Switching Underlay NW
Require dynamically managed VLAN configuration per port in the physical switch fabric
Bare Metal Cloud
SmartNIC implemented VIM NW separation enabling infrastructure acceleration for high performance workloads
Server
V/CNF
Controlled by SW Infrastructure Management
Virtual Switch
vSwitch
VLAN
SDNo
VIM
Controlled by HW Infrastructure Management
Smart NIC
VNI Range x
VTEP
SDNu
VxLAN Overlay NW
Physical Switch Fabric
Switching Underlay NW
Dynamically managed VLAN configuration
Statically provisioned switch fabric at physical data center deployment

SDN Overlay and SDN Underlay concepts, layering and relationships

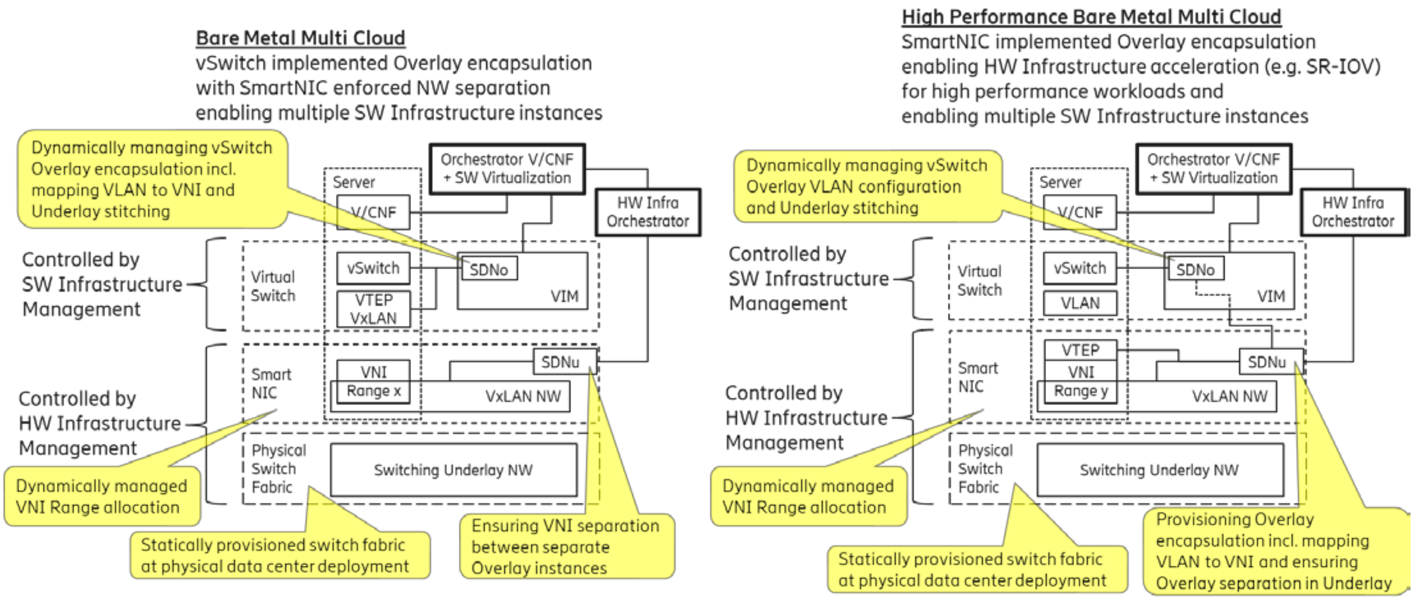
An SDN Overlay controller (here denoted SDNo) is responsible for managing the SW Virtualization Layer virtual switching that manages the Overlay Network switching and encapsulation and mapping onto the Underlay Networks.

In cases where the V/CNF bypasses the SW Virtualization Layer virtual switching e.g. high performance applications using SR-IOV, there is a need for the HW Infrastructure Layer to perform the encapsulation and mapping onto the Underlay Networking. This is controlled by the SDN Underlay controller (SDNu) that is authoritative and ensures separation of the shared Underlay Networking resources.

SDNo controllers can request Underlay encapsulation and mapping to be done by signaling to an SDNu controller. There are however today no standardized way for this signaling and by that there is a missing reference point and API description in this architecture.

For deployments with multiple SW Virtualization instances sharing the same Underlay Networking resources, it is the SDNu responsibility to ensure separation of the shared Underlay resources and to set up appropriate enforcements in the Underlay Networking forwarding plane since each SW Virtualization instance or V/CNF that bypasses its local virtual switching instance cannot be trusted. Any fault or misconfiguration in such instances could potentially destroy all networking in the shared Underlay Networking.

Two use case examples with both SDNo and SDNu controllers depicting a normal virtual switch encapsulating SW Virtualization Infrastructure instance and another high performance oriented SW Virtualization Infrastructure instance (e.g. using SR-IOV) are described in the figure below. The example is showing how the encapsulation and mapping could be done in the virtual switch or in a SmartNIC on top of a statically provisioned underlay switching fabric, but another example could also have been depicted with the SDNu controlling the underlay switching fabric without usage of SmartNICs.



Bare Metal Multi Cloud
vSwitch implemented Overlay encapsulation
with SmartNIC enforced NW separation
enabling multiple SW Infrastructure instances

Dynamically managing vSwitch Overlay encapsulation incl. mapping VLAN to VNI and Underlay stitching

Dynamically managed VNI Range allocation

Statically provisioned switch fabric at physical data center deployment

Ensuring VNI separation between separate Overlay instances

High Performance Bare Metal Multi Cloud
SmartNIC implemented Overlay encapsulation
enabling HW Infrastructure acceleration (e.g. SR-IOV)
for high performance workloads and
enabling multiple SW Infrastructure instances

Dynamically managing vSwitch Overlay VLAN configuration and Underlay stitching

Dynamically managed VNI Range allocation

Statically provisioned switch fabric at physical data center deployment

Provisioning Overlay encapsulation incl. mapping VLAN to VNI and ensuring Overlay separation in Underlay

Programmable Networking Fabric

The concept of a Programmable Networking Fabric pertains to the ability to have an effective forwarding pipeline (a.k.a. forwarding plane) that can be programmed and/or configured without any risk of disruption to the shared Underlay Networking that is involved with the reprogramming for the specific efficiency increase.

The forwarding plane is distributed by nature and must be possible to implement both in switch elements and on SmartNICs (managed outside the reach of host SW) that both can be managed from a logically centralized control plane.

The logically centralized control plane is the foundation for the authoritative separation in between different SW Virtualization domains or Bare Metal Network Function applications that are regarded as untrusted both from the shared layers and each other.

Although the control plane is logically centralized, scaling and control latency concerns must allow the actual implementation of the control plane to be distributed when required.

All VNF, CNF and SW Virtualization acceleration or specific support functionality that could be programmed in the forwarding plane must be confined to the well separated sections or stages of any shared Underlay Networking. A practical example could be a SW Virtualization instance or VNF/CNF that control a NIC/SmartNIC where the Underlay Networking (switching fabric) ensures the separation in the same way as it is done for SR-IOV and PCI-PT cases today.

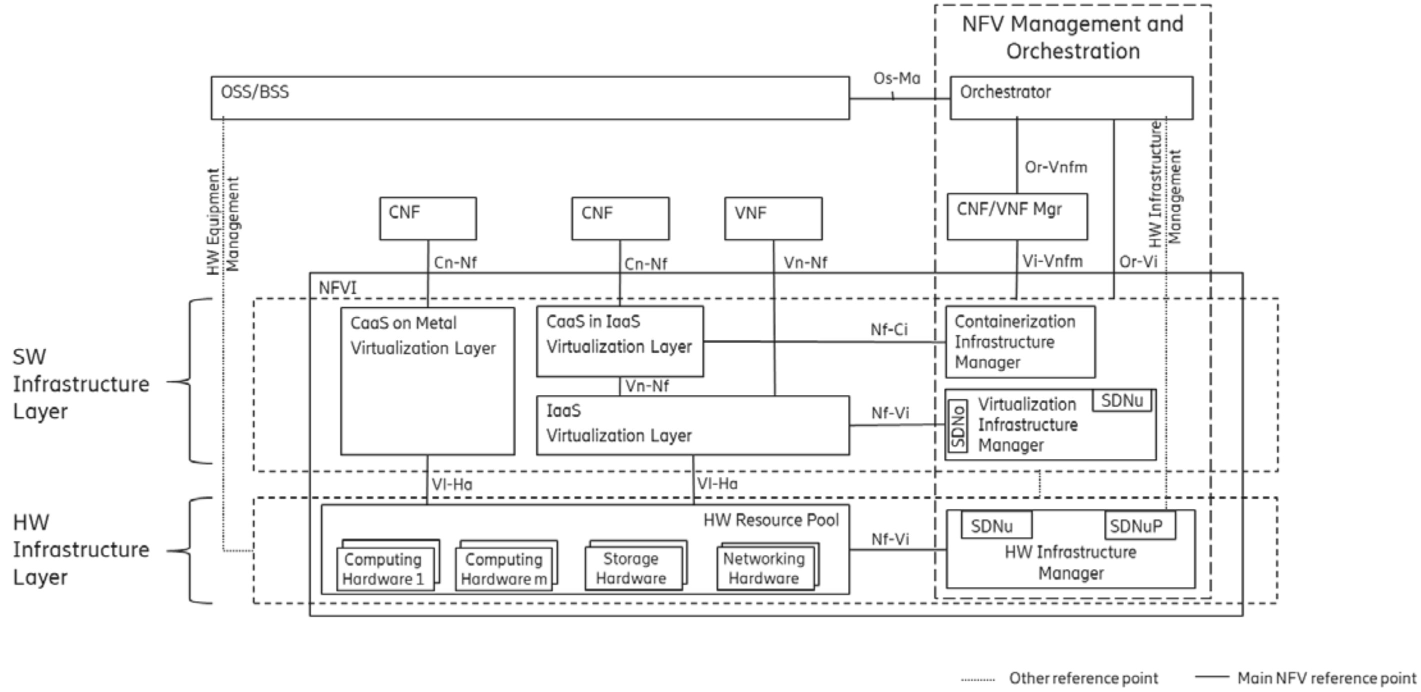
The nature of a shared Underlay Network that shall ensure separation and be robust is that all code in the forwarding plane and in the control plane must be under the scrutiny and life cycle management of the HW Infrastructure Layer.

This also imply that programmable forwarding functions in a Programmable Networking Fabric are shared resources and by that will have to get standardized interfaces over time to be useful for multi-vendor architectures such as ETSI NFV. Example of such future extensions of shared functionality implemented by a Programmable Networking Fabric could be L3 as a Service, Firewall as a Service and Load Balancing as a Service.

Note: Appliance-like applications that fully own its infrastructure layers (share nothing) could manage and utilize a Programmable Networking Fabric in many ways, but that is not a Cloud implementation and falls outside the use cases for these specifications.

Networking Reference Model

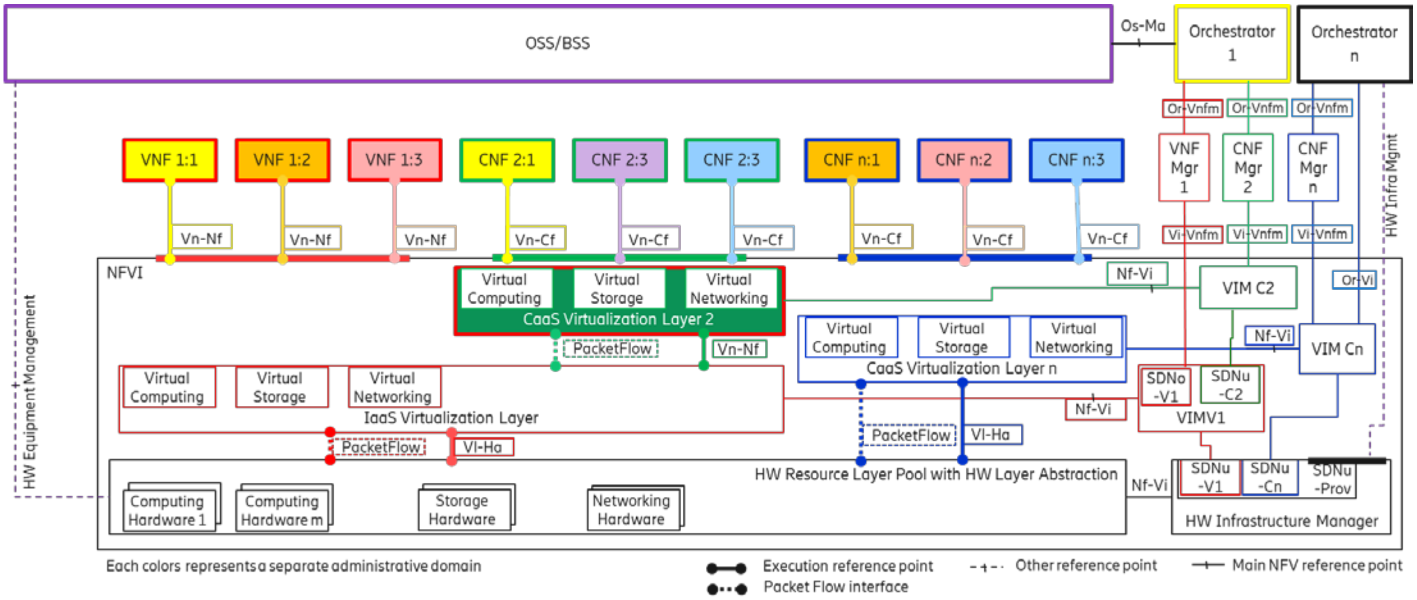
The Networking Reference Model depicted in the figure below is based on the ETSI NFV model and has a strict separation of the HW Infrastructure and SW Infrastructure Layers in NFVI. It enables multiple well separated simultaneous SW Virtualization domains allowing a mix of CaaS on Metal, CaaS on IaaS and IaaS on a shared HW infrastructure.



Deployment example of a Networking Reference Model

A Networking Reference Model deployment example is depicted in the figure below to demonstrate the mapping to ETSI NFV reference points with additions of packet flows through the infrastructure layers and some other needed reference points. The example illustrates individual responsibilities of a complex organization with multiple separated administrative domains here represented with separate colors.

The example is a rather common example scenario from operators that is during a rather long period of migration of some of the network functions from the predominant VNF / IaaS environments of today into a selection of CNF / CaaS on bare Metal and other on IaaS.



OSS/BSS
Os-Ma
Orchestrator 1
Orchestrator n
Or-Vnfm
Or-Vnfm
Or-Vnfm
VNF 1:1
VNF 1:2
VNF 1:3
CNF 2:1
CNF 2:3
CNF 2:3
CNF n:1
CNF n:2
CNF n:3
VNF Mgr 1
CNF Mgr 2
CNF Mgr n
HW Infra Mgmt
Vn-Nf
Vn-Nf
Vn-Nf
Vn-Cf
Vn-Cf
Vn-Cf
Vn-Cf
Vn-Cf
Vn-Cf
Vi-Vnfm
Vi-Vnfm
Vi-Vnfm
NFVI
Virtual Computing
Virtual Storage
Virtual Networking
CaaS Virtualization Layer 2
Nf-Vi
VIM C2
Or-Vi
PacketFlow
Vn-Nf
Virtual Computing
Virtual Storage
Virtual Networking
CaaS Virtualization Layer n
Nf-Vi
VIM Cn
HW Equipment Management
Virtual Computing
Virtual Storage
Virtual Networking
IaaS Virtualization Layer
SDNo -V1
SDNu -C2
Nf-Vi
VIMV1
PacketFlow
VI-Ha
PacketFlow
VI-Ha
HW Resource Layer Pool with HW Layer Abstraction
Nf-Vi
SDNu -V1
SDNu -Cn
SDNu -Prov
Computing Hardware 1
Computing Hardware m
Storage Hardware
Networking Hardware
HW Infrastructure Manager
Each colors represents a separate administrative domain
Execution reference point
Packet Flow interface
Other reference point
Main NFV reference point