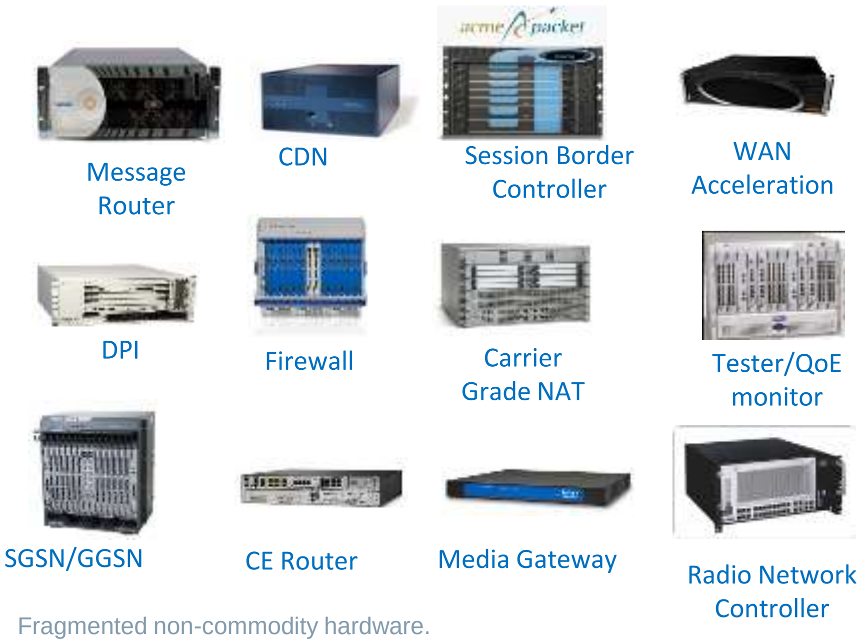
Distributed NFV & OpenStack: Challenges and potential solutions

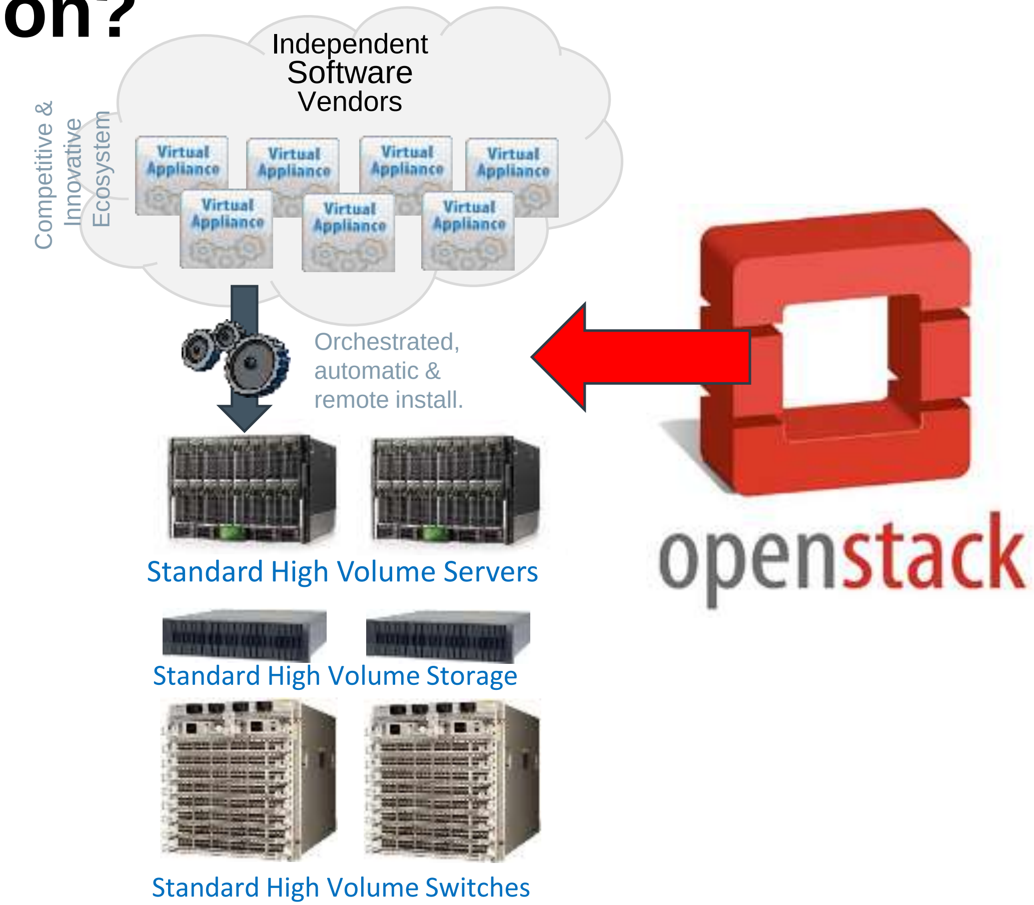
Tariq Khan, HPE.
Adrian Hoban, Intel.
Prithiv Mohan, Intel.
Arun Thulasi, HPE.
Peter Willis, BT.

What is Network Functions Virtualisation?

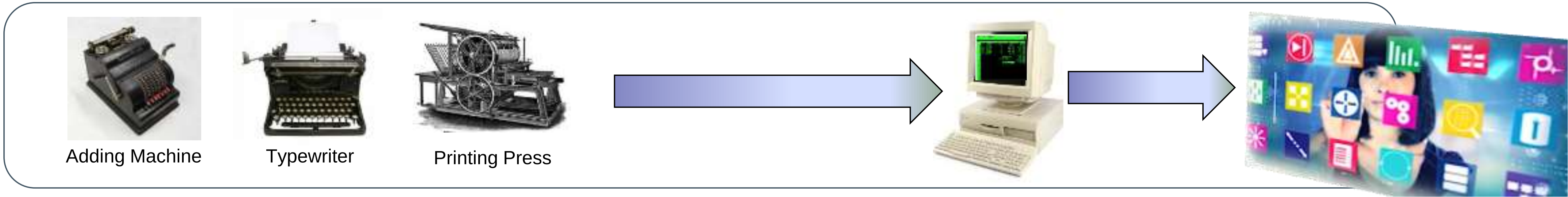
Classical Network Appliance Approach



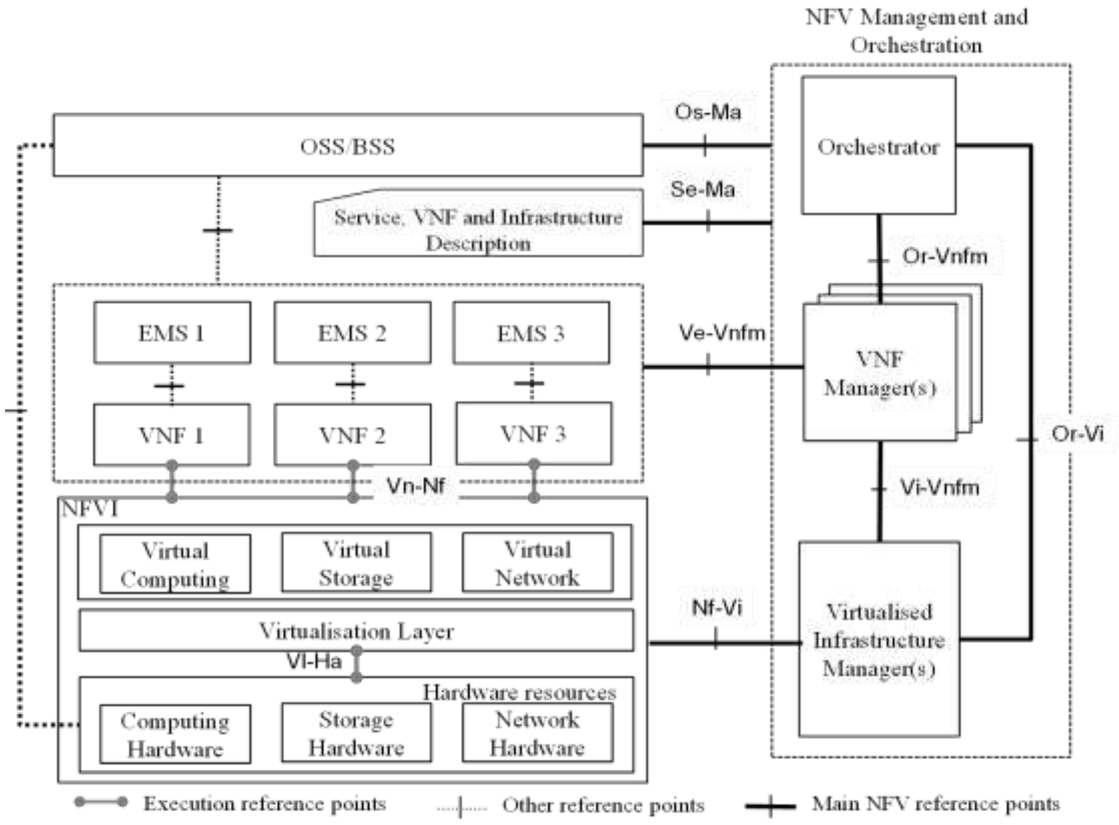
Fragmented non-commodity hardware.
Physical install per appliance per site.
Hardware development large barrier to entry for new vendors constraining innovation & competition.


Independent Software Vendors
Competitive & Innovative Ecosystem
Virtual Appliance
Orchestrated, automatic & remote install.
Standard High Volume Servers
Standard High Volume Storage
Standard High Volume Switches
openstack

Network *functions* Virtualisation Approach



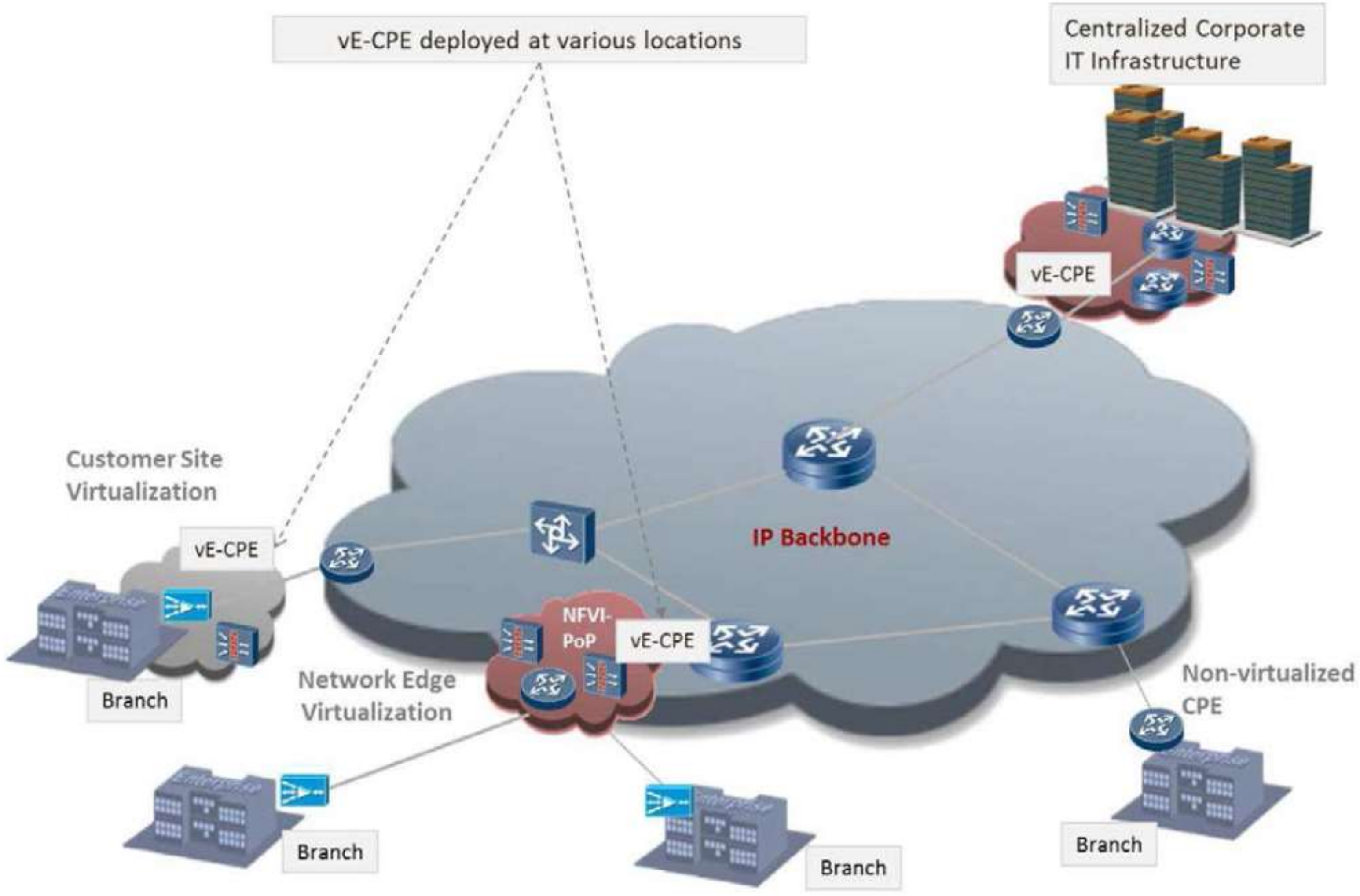
Why use OpenStack for NFV?



ETSI NFV ISG reference architectural framework

- The Virtualised Infrastructure Manager is part of the ETSI NFV Industry Specification Group's Architecture.
  - See http://www.etsi.org/deliver/etsi_gs/NFV/001_099/002/01.02.01_60/gs_NFV002v010201p.pdf
- The VIM was always assumed to reuse existing IT solutions e.g: OpenStack etc. because NFV and Cloud were seen as similar but not identical.
  - The ISG did not want to reinvent the VIM.
- Many assumed OpenStack core projects as the VIM because of its development momentum.
- But are the differences between NFV and Cloud too large a gap for OpenStack to bridge?

What is Distributed NFV a.k.a Virtual Enterprise CPE (vE-CPE)



vE-CPE deployed at various locations
Centralized Corporate IT Infrastructure
vE-CPE
Customer Site Virtualization
vE-CPE
IP Backbone
NFVI-PoP
vE-CPE
Branch
Network Edge Virtualization
Non-virtualized CPE
Branch
Branch
Branch

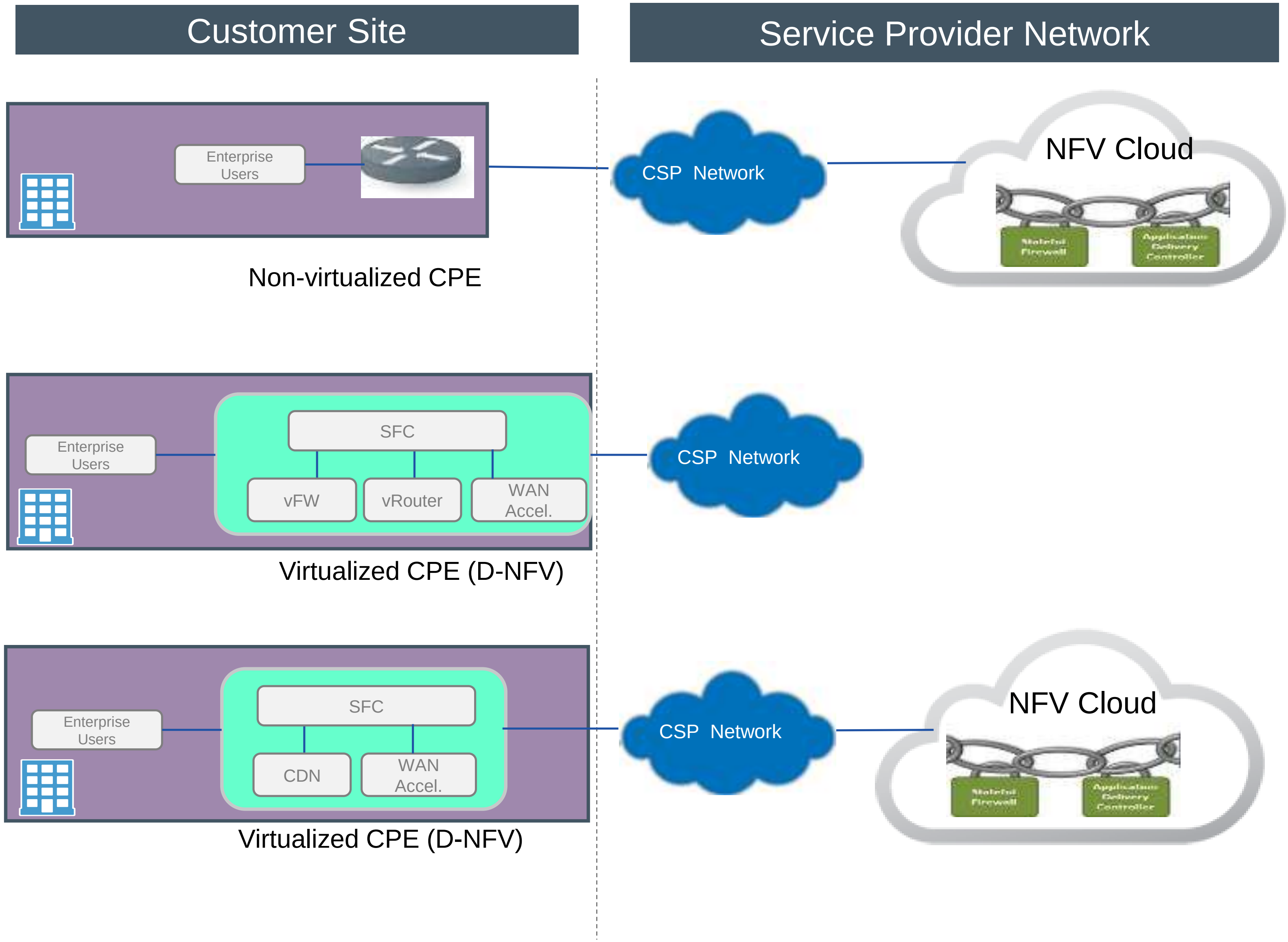
– D-NFV implements network functions on servers at the customer premises.

– Example functions that need to be implemented on premise are:

1. WAN acceleration
2. Security functions dependent on policy
3. Upstream QoS
4. Protection Switching
5. Instrumentation
6. Router dependent on access link

Source: ETSI Use case http://www.etsi.org/deliver/etsi_gs/nfv/001_099/001/01.01.01_60/gs_nfv001v010101p.pdf

Deployment Options



- Centralized/Hosted Model

Simple CPE.
All services hosted in the NFV cloud
Need CPE ⇔ NFV cloud tunnel

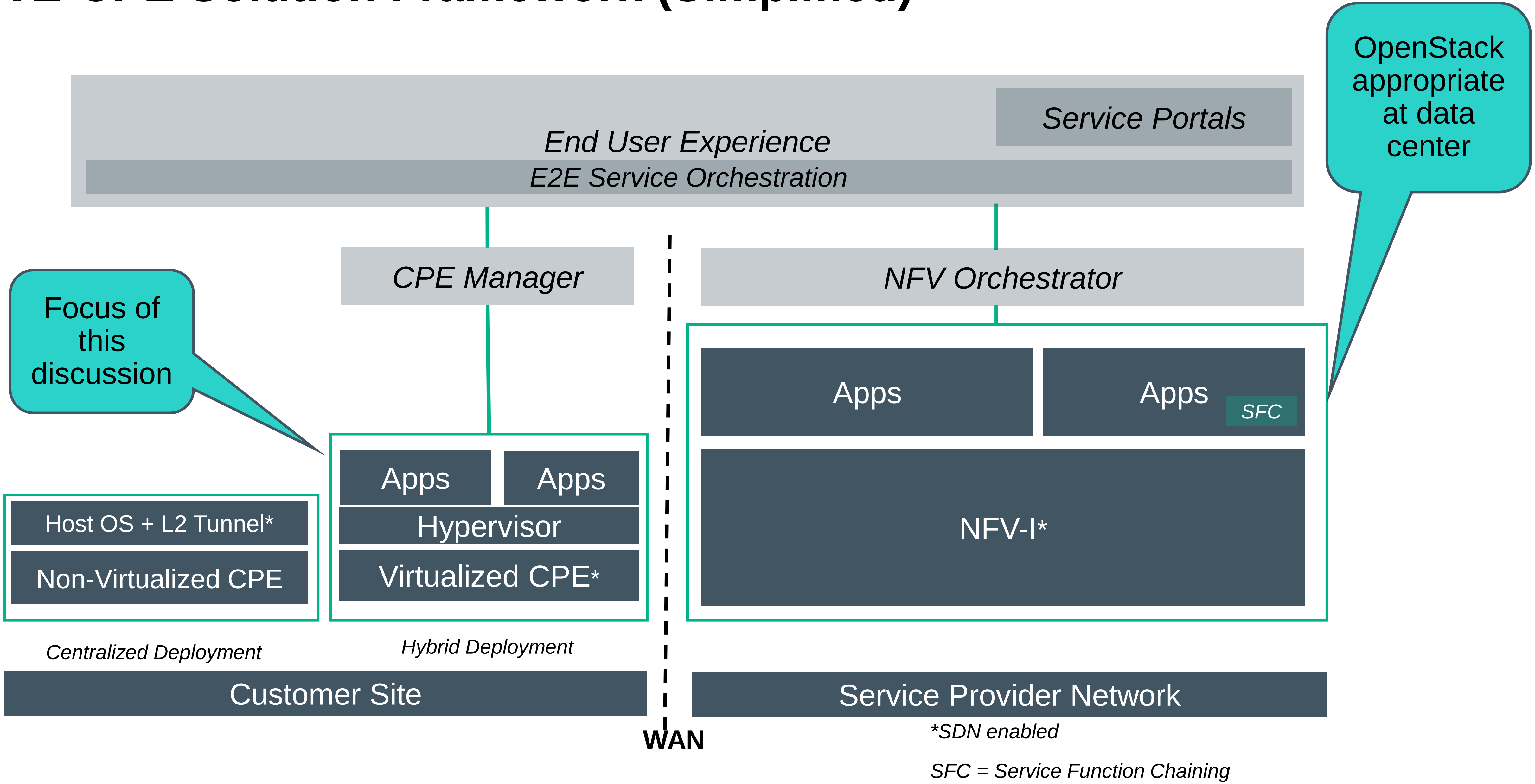
Distributed/De-centralized Model

CPE supports virtualization.
All services placed at customer site.

Hybrid/Mixed Model

Services placed both at customer site and in NFV Cloud
Flexible model

vE-CPE Solution Framework (Simplified)



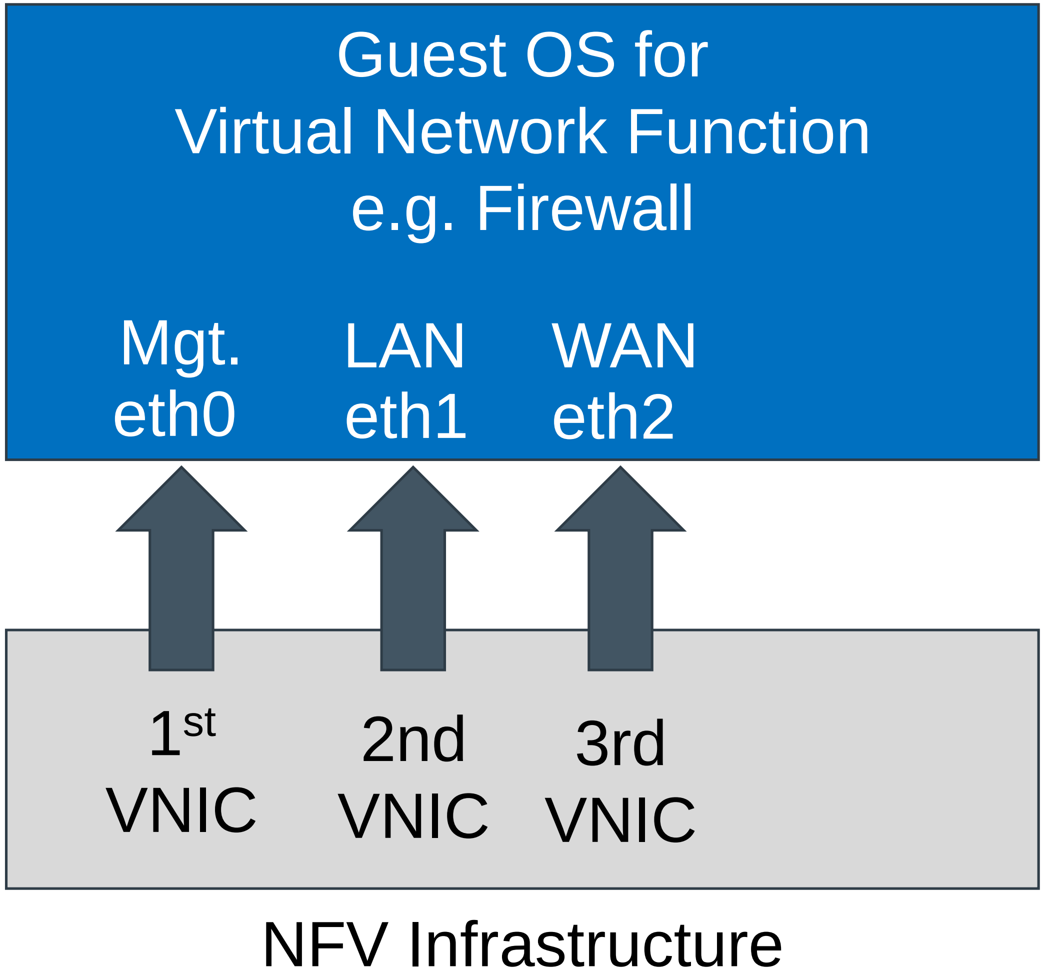
6 Top Challenges for Using OpenStack For D-NFV

1. Binding Virtual Network Interface Cards to the Virtual Network Function
2. Service chain modification
3. Securing OpenStack over the Internet
4. Scalability of the controller(s)
5. Start-up Storms (Or Stampedes)
6. Backwards compatibility between releases

Challenge 1: Binding Virtual Network Interface Cards to the Virtual Network Function

- VNFs typically number interfaces according to order they are connected to the guest VM.
  - VNICs are connected in sequence.

- Challenge 1A: How do we verify the correct VNF interface has been connected to the correct VNIC?



Guest OS for
Virtual Network Function
e.g. Firewall

Mgt.
eth0

LAN
eth1

WAN
eth2

1st
VNIC

2nd
VNIC

3rd
VNIC

NFV Infrastructure

Background: Consistent Device Naming (CDN) in Linux

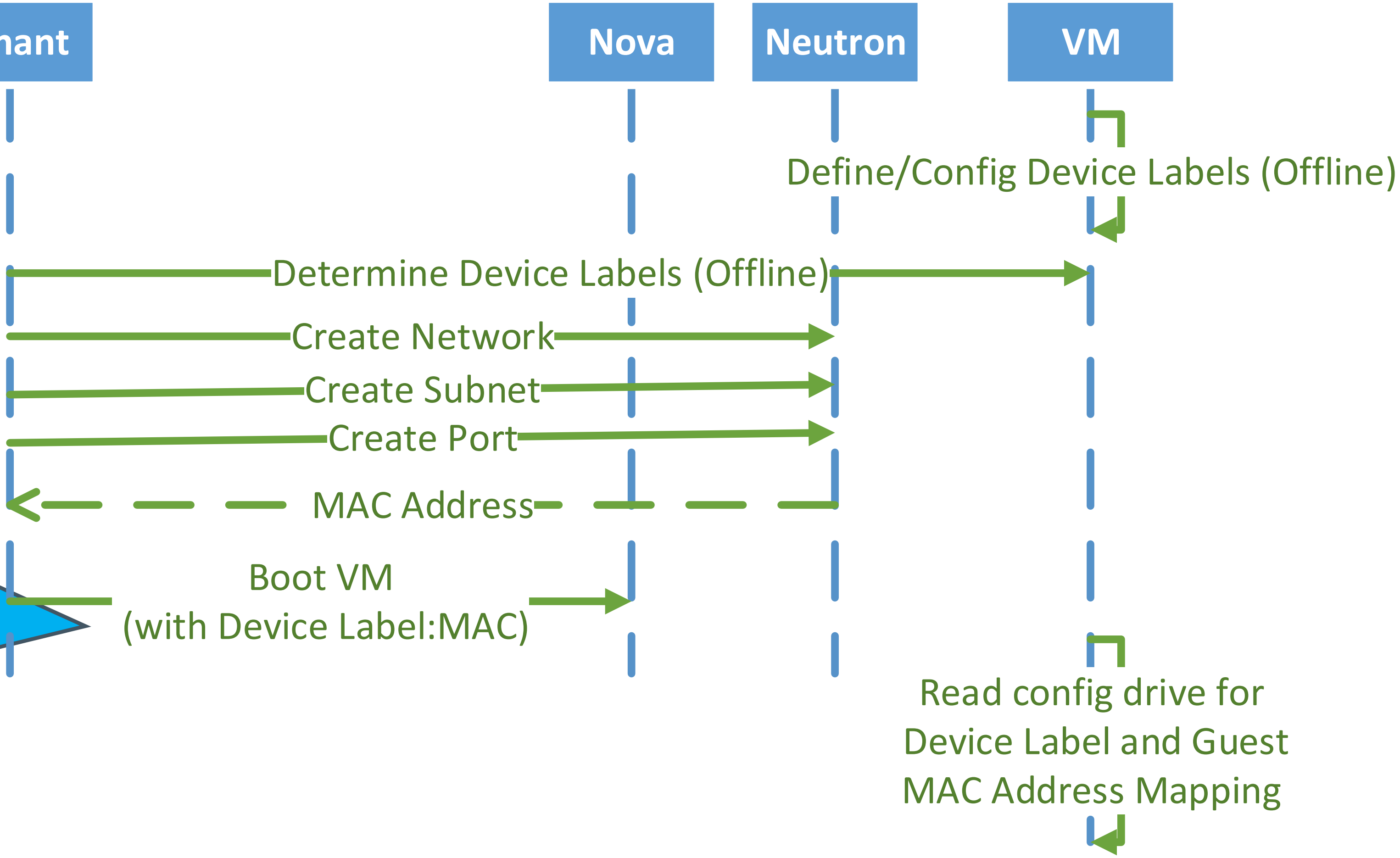
– CDN Convention:

- – Embedded LAN-on-Motherboard: `em[123…]`, e.g. `em2`
- – PCIe Network Card: `p<slot>p<Ethernet port>`, e.g. `p4p0`
- – Virtual Function: `p<slot>p<Ethernet port>_<virtual interface>`, e.g. `p4p0_2`
- – Note: System admins may overwrite this naming convention.
  - – E.g. in Fedora in `/etc/udev/rules.d/70-persistent-net.rules`

– CDN must be enabled with the `biosdevname=1` kernel boot setting (usually default)

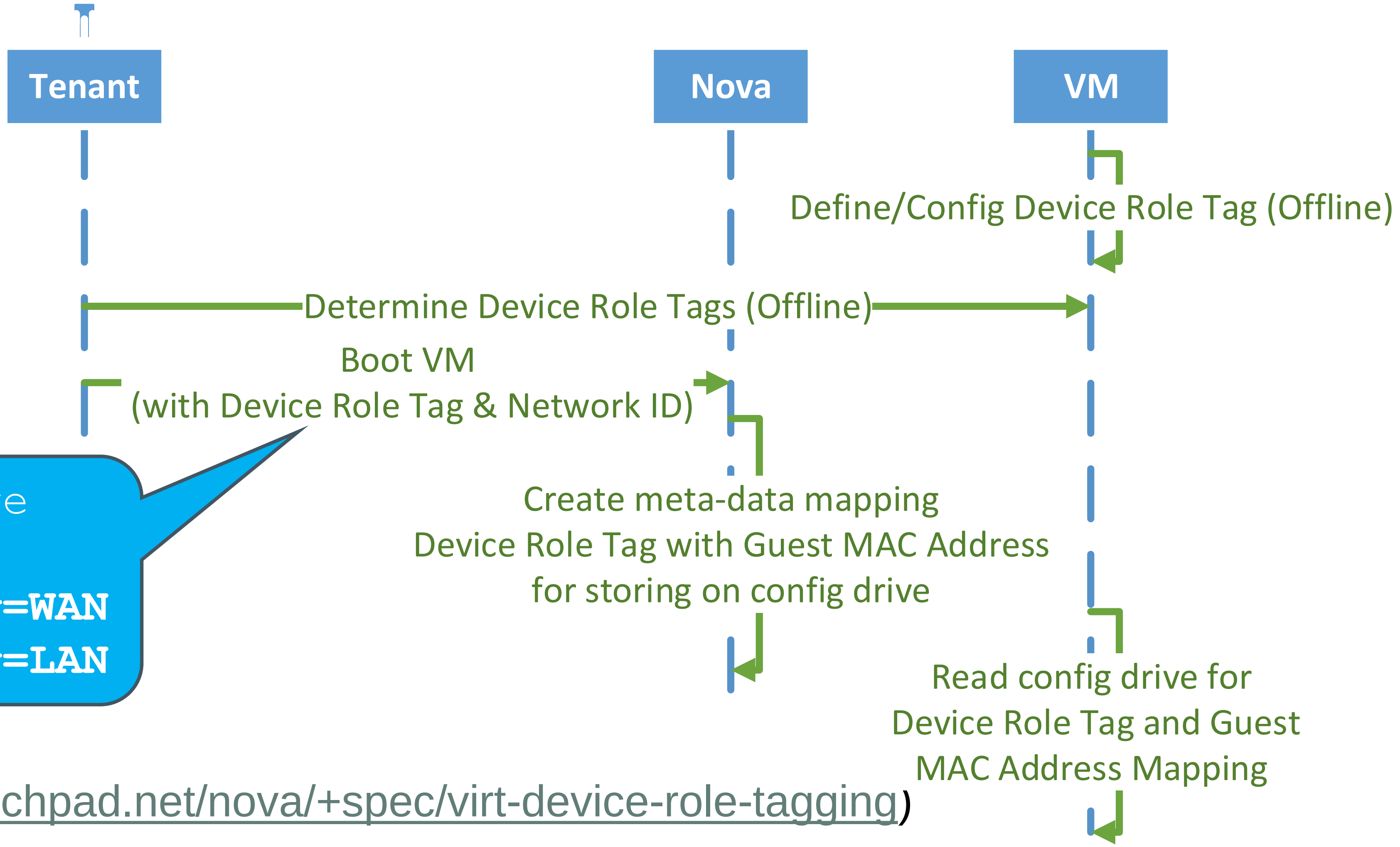
Recommendation: Do not depend on this capability to resolve challenge 1a

Short Term Solution 1a: Nova Boot With Meta-data Info (Proposal to support configuration drive enabled VNFs)



```
nova boot --image
NFVimage --flavor
m1.nfv_large --nic
port-id=12345
--meta <Device
Label>:<MAC Address>
```

Longer Term Solution 1a: Virtual Guest Device Role Tagging (Proposal to support configuration drive enabled VNFs)


Tenant
Nova
VM
Define/Config Device Role Tag (Offline)
Determine Device Role Tags (Offline)
Boot VM
(with Device Role Tag & Network ID)
Create meta-data mapping
Device Role Tag with Guest MAC Address
for storing on config drive
Read config drive for
Device Role Tag and Guest
MAC Address Mapping

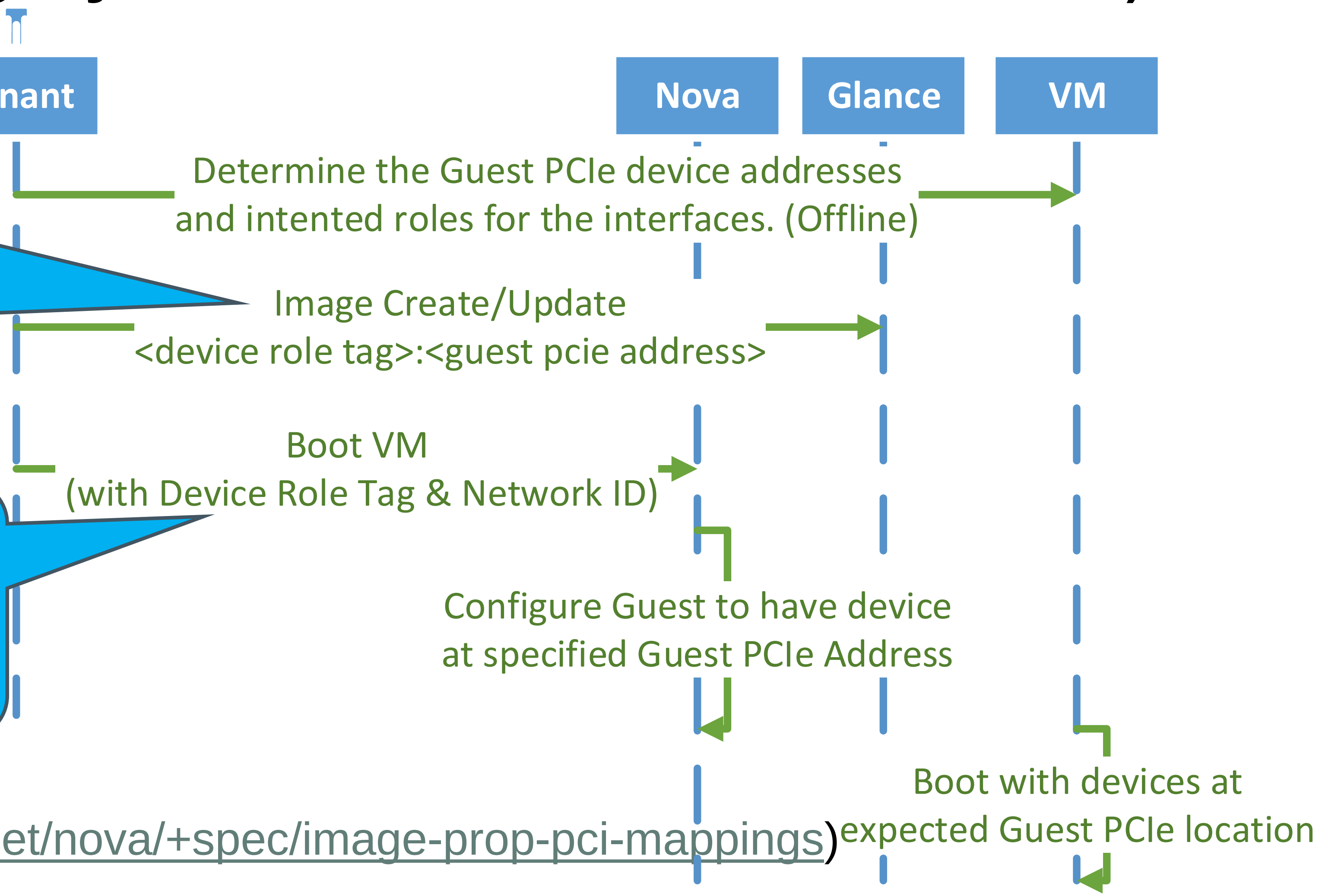
```
nova boot --image NFVimage
  --flavor m1.nfv_large
  --nic net-id=12345, tag=WAN
  --nic net-id=56789, tag=LAN
```

- Blueprint: https://blueprints.launchpad.net/nova/+spec/virt-device-role-tagging)

Longer Term Solution 1a: Virtual Guest Device Role Tagging (Proposal to support “legacy” VNFs with additional extension)

```
glance image-update
--name NFVimage
--property
hw_pci_mappings=WAN:<00:83:00:0>
--property
hw_pci_mappings=LAN:<00:83:01:0>
```

```
nova boot --image NFVimage
  --flavor m1.nfv_large
  --nic net-id=12345, tag=WAN
  --nic net-id=56789, tag=LAN
```


Tenant
Nova
Glance
VM
Determine the Guest PCIe device addresses and intented roles for the interfaces. (Offline)
Image Create/Update <device role tag>:<guest pcie address>
Boot VM (with Device Role Tag & Network ID)
Configure Guest to have device at specified Guest PCIe Address
Boot with devices at expected Guest PCIe location

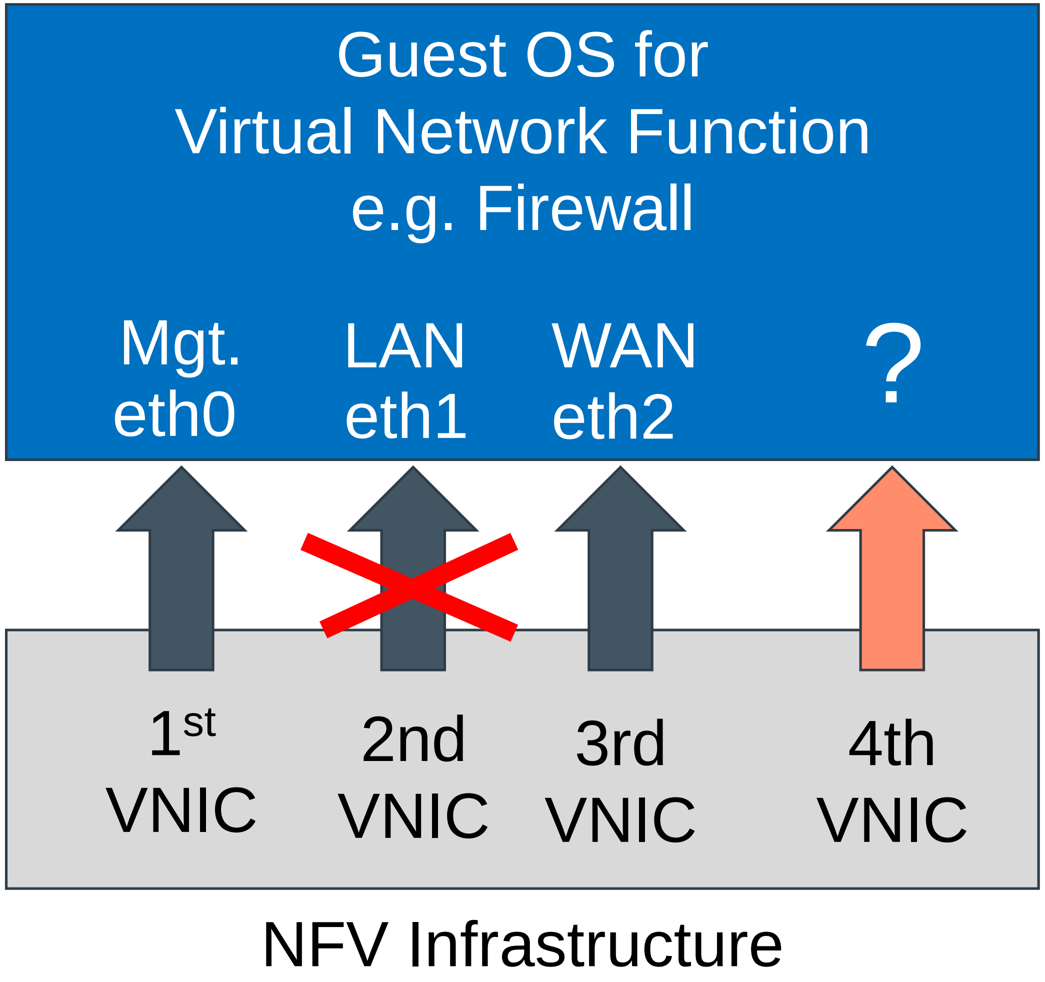
- Blueprint:https://blueprints.launchpad.net/nova/+spec/image-prop-pci-mappings)

Challenge 1: Binding Virtual Network Interface Cards to the Virtual Network Function

- Challenge 1B: What happens when an interface is disconnected and reconnected?

- In practice three different behaviours have been observed:

1. VNIC reconnects as eth1
2. VNIC reconnects as eth3
3. VNF locks-up



Solutions 1b: Interface Disconnect/Reconnect

Short term:

- Primary reason for the disconnect event related to topology changes when leveraging the neutron network chaining model.
  - Moving to an SFC model will remove the source of these disconnect/reconnect events.
- Leverage CDN as best effort to handle disconnect/reconnect events due to other triggers.

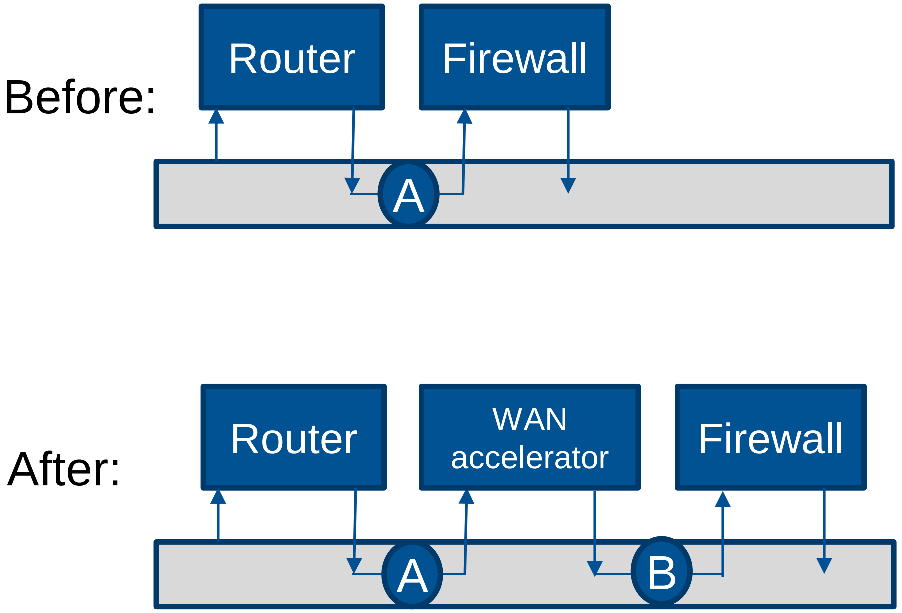
Longer term:

- Collaborate with the VNF vendor community to better handle disconnect/reconnect events.

Challenge 2: Service Chain Modification

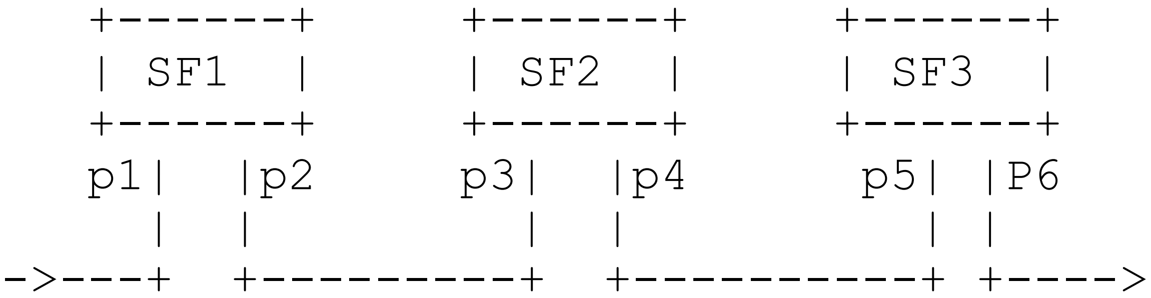
When chaining services via dedicated Neutron Networks per service:

- OpenStack has no primitives to reconnect interface on Firewall from A to B.
- Can only be done by deleting the interface on the firewall and reconnecting which leads to ambiguity.
- Or provision a total new service chain from scratch which causes >5 minutes outage.



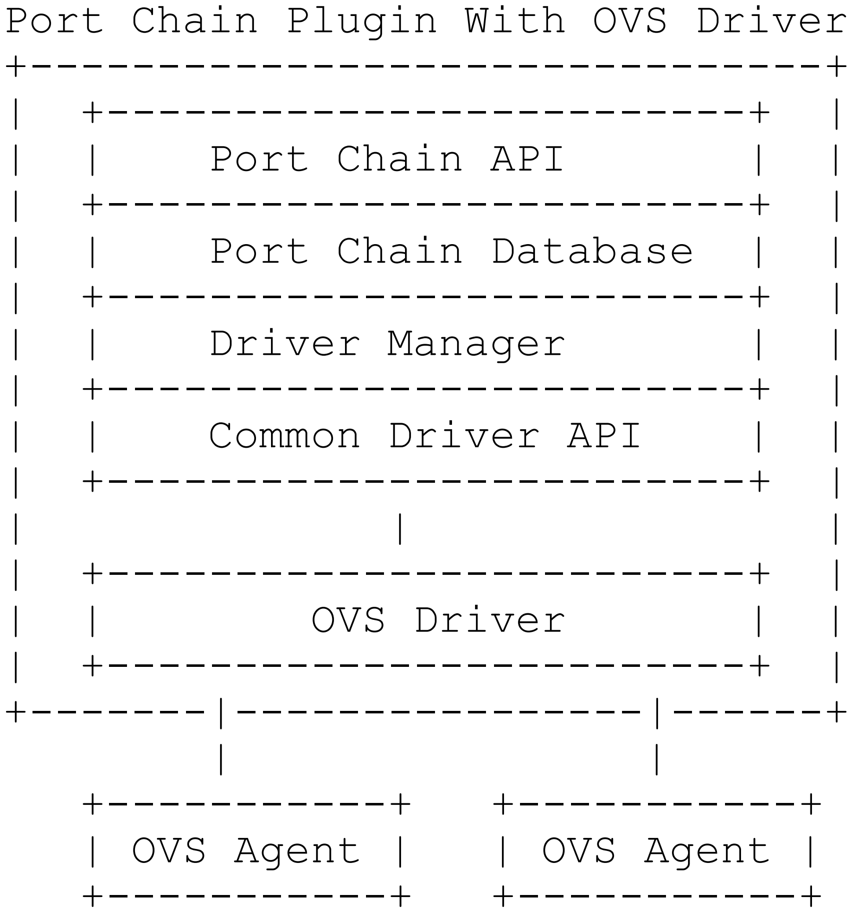
Short-Term Solution 2: OpenStack networking-sfc

- OpenStack networking-sfc project provides a “Port Group” based mechanism to address service chaining use cases.
- Relatively easy to consume API



```
   +-------+      +-------+      +-------+
   |  SF1  |      |  SF2  |      |  SF3  |
   +-------+      +-------+      +-------+
    p1|  |p2      p3|  |p4        p5| |P6
      |  |          |  |            | |
->----+  +----------+  +------------+ +---->
```

Figure from: http://docs.openstack.org/developer/networking-sfc/api.html

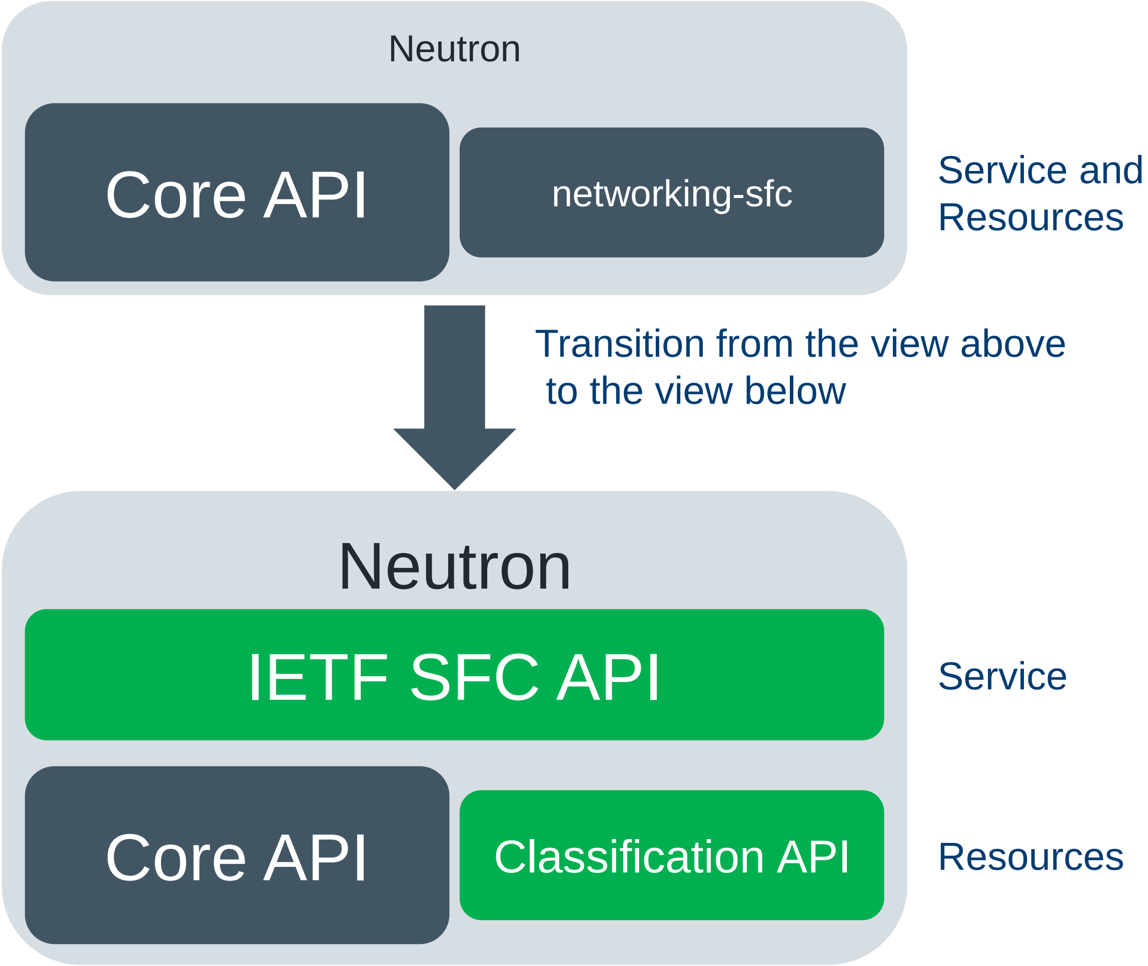


```
Port Chain Plugin With OVS Driver
+--------------------------------+
|  +--------------------------+  |
|  |    Port Chain API        |  |
|  +--------------------------+  |
|  |    Port Chain Database   |  |
|  +--------------------------+  |
|  |    Driver Manager        |  |
|  +--------------------------+  |
|  |    Common Driver API     |  |
|  +--------------------------+  |
|               |                |
|  +--------------------------+  |
|  |        OVS Driver        |  |
|  +--------------------------+  |
+-------|----------------|-------+
        |                |
  +-----------+   +-----------+
  | OVS Agent |   | OVS Agent |
  +-----------+   +-----------+
```

Figure from: http://docs.openstack.org/developer/networking-sfc/system_design%20and_workflow.html#system-architecture

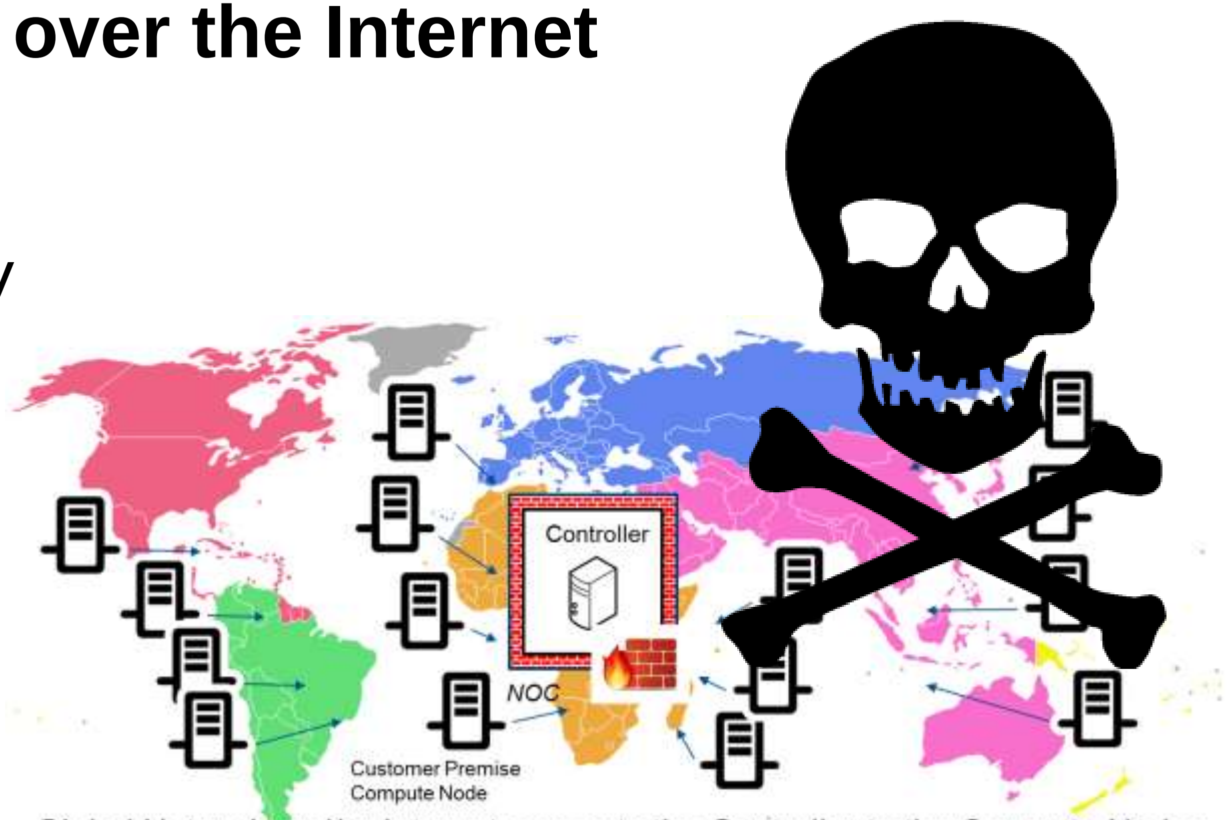
Longer Term Solution 2: OpenStack networking-sfc v2

- Initiative started looking at advancing the SFC capabilities in Neutron
  - Follow mail thread with subject: “A standards-compliant SFC API”
- Goal to address topics identified (at time of writing) in current approach such as:
  - Coupled Resources and Service management views
  - Lack of IETF compliance
  - Hidden Service Path information (cannot be integrated with an NFVO)
  - No support for meta-data
  - No SFC encapsulation (IETF NSH or other);
  - No instance selection policies;
  - Current API and respective implementation have an implicit SFC Proxy around the approximation of SFF, that cannot be disabled;
  - Limited traffic classification;



Challenge 3: Securing OpenStack over the Internet

- –BT Research connected a compute node over the Internet to a controller in their NFV Lab.
- –Over 500 pin holes had to be opened in the firewall to allow this to work
  - –Includes ports for VNC and SSH for CLIs.
- –Firewall had to be reconfigured every time the compute node’s dynamic IP address changed.
  - –Which it did several times during testing.
- –It is a realistic scenario for the vBranch.
- –OpenStack’s design presents too many attack vectors.



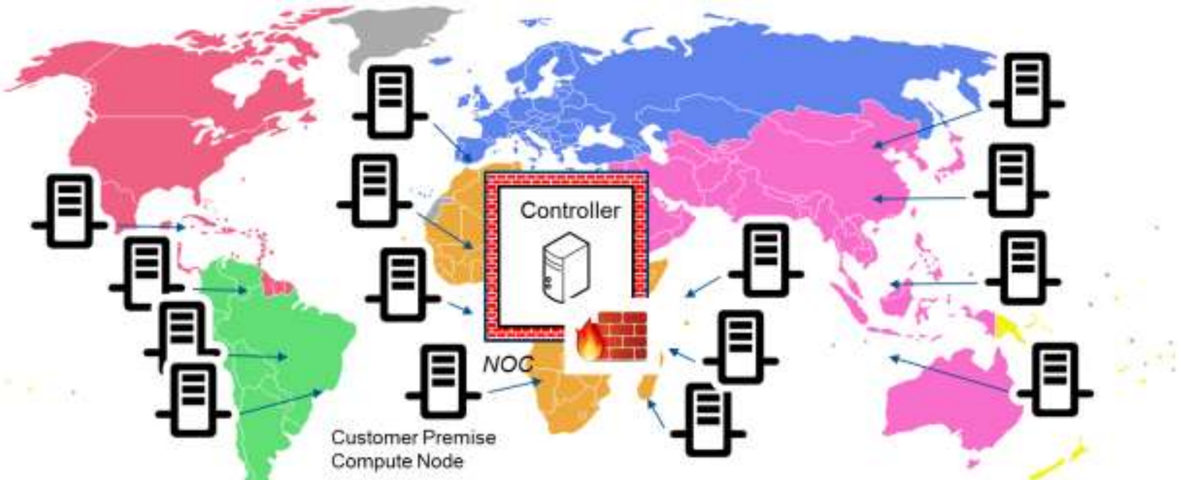
Global Network and/or Internet connects the Controller to the Compute Nodes

Solutions to Challenge 3: Securing OpenStack over the Internet

- Short Term:
  - Tunnel all the OpenStack control traffic into a VPN service (managed outside of OpenStack control)
    - IPsec most obvious & portable technology to use
    - Need to be careful where IPsec is terminated on the CPE w.r.t. firewalls, hypervisors and agents, otherwise might still end up exposing agents to the Internet.
- Medium to Longer Term:
  - Consider alternative deployment architectures such as having federated local OpenStack services and remote shared services.

Challenge 4: Scalability of the OpenStack Controller

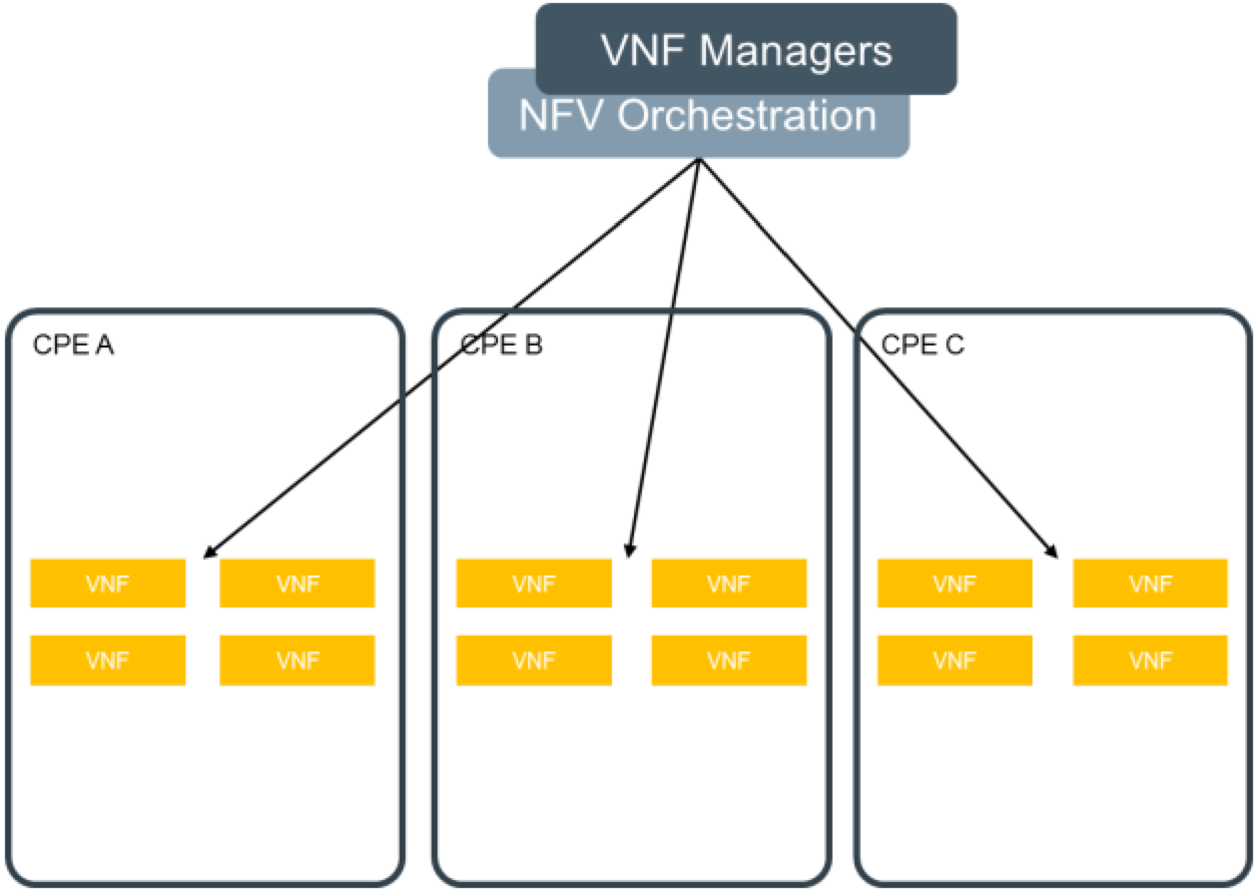
- –How many compute nodes can be connected to a single controller?
  - –<500!
- –How is the scalability of a controller tested?
- –Should controllers be regionalised?
  - –What size of region?
    - –Continent, Country, Town, Branch?
    - –Or per customer?



Global Network and/or Internet connects the Controller to the Compute Nodes

Short Term Solution 4: No OpenStack control for edge

- Don’t use OpenStack to manage devices at customer premise
- Use a lightweight hyper-visor manager (libvirt/virt-manager)
- Pro:
  - scale limited by orchestration & management software; smallest compute nodes possible
- Con:
  - can’t leverage OpenStack virtualization management capabilities, all management using NFVO or other software
  - Requires new functional element in the architecture – CPE Manager



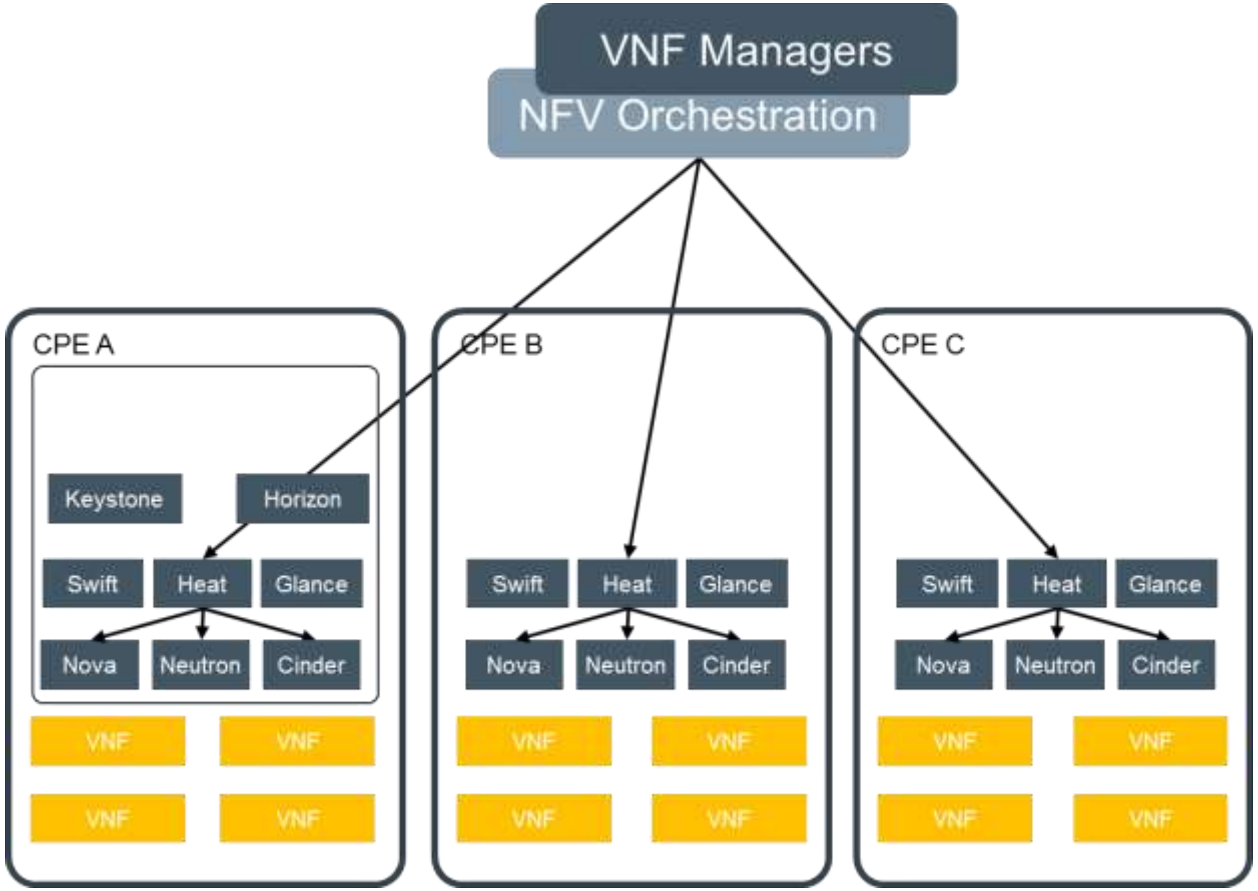
Medium Term Solution 4: Hybrid Compute & Control node at the edge

–Lightweight control plane (federated OpenStack with local (e.g. nova, neutron & cinder) and remote shared services)

–Use OpenStack Kolla to deploy containerized light weight controller services on remote sites

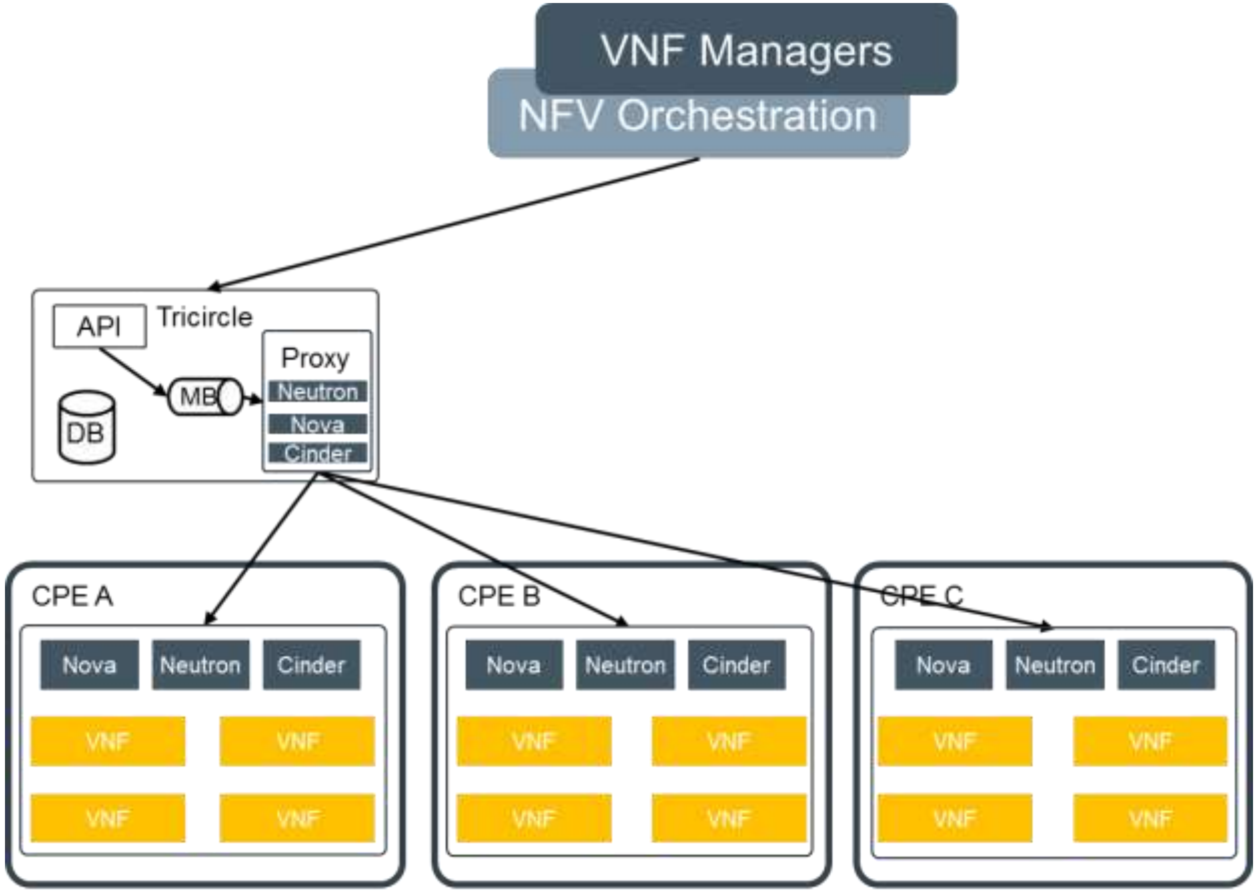
–Pro: distributed scale, common API endpoints, smaller compute nodes

–Con: requires careful design, implementation and operational processes



Long Term Solution 4: Hybrid control & compute at the edge with single API endpoint

- –Combination of OpenStack TriCircle & Kolla
- –Pro: distributed scale, single API endpoint
- –Con: need community engagement to move it forward



* Only showing nova, neutron & cinder for simplicity

Challenge 5: Start-up Storms (Or Stampedes)

What Happens

– Disaster scenario at customer data center

– Upon recovery and power-up, all OpenStack clients reaching back to controllers for OpenStack information

– VMs/Applications trying to reach neighbors and re-establish connectivity

Why Is It Important

– OpenStack controller services overwhelmed with client requests

– Centralized data center hosted VNFs overwhelmed with connection re-establishment requests

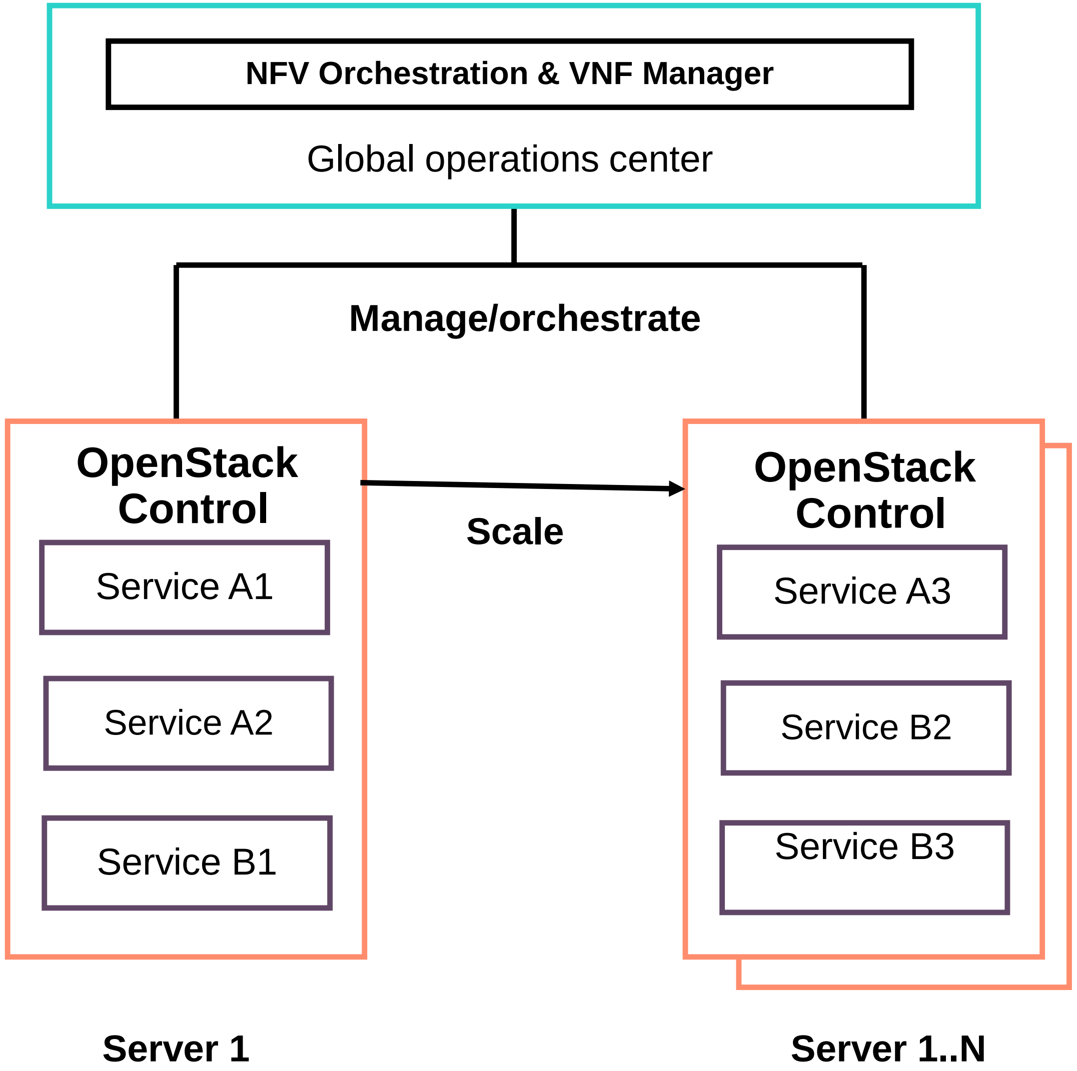
– Local network jammed due to ARP storms

[illegible] (RF) © www.visualphotos.com

Hewlett Packard Enterprise

Short term solution 5: Deployment Driven

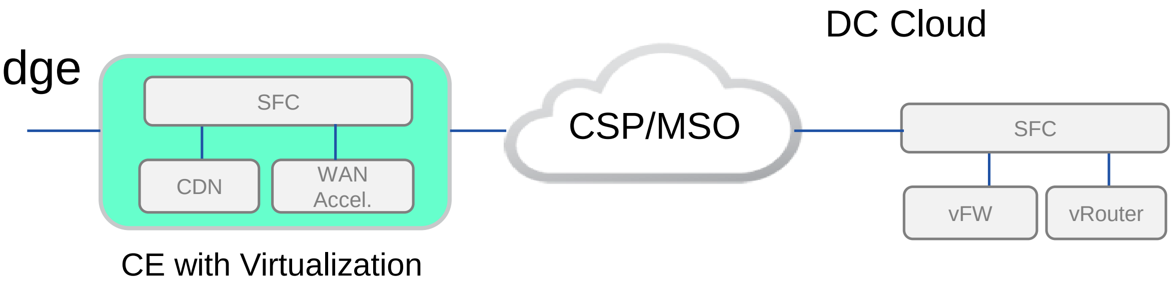
- **Dynamic scaling of Control instances**
  - Services that are impacted automatically scale based on load
  - Load continues to exist on the networking equipment
- **Networking Protocol advantages**
  - ARP storms a potential situation in a startup storm
  - IPv6 and Neighbor Discovery to eliminate ARP storms.



Hewlett Packard Enterprise

Long Term Solution 5: Platform & VNF Driven

- **Keep Storms Local: OpenStack L2/L3 services run at the edge**
  - Extend neutron L2/L3 services scaled up to run at the edge
- **Keep Storms Local: Condensed OpenStack services at the edge.**
  - Services that require local networking to run “headless”
  - “Headless” services that are connected services that report back to the central site
- **Hybrid vCPE deployment model to run VNFs on Edge devices.**
  - OpenStack services being impacted by load is part of the challenge. VNFs being impacted is larger challenge.
  - Hybrid models allow localized VNF services to run at the edge instead of NG-POP



Hybrid Edge/Cloud Model
Services placed both in CPE/DC cloud
Allows transitioning from edge to cloud
Flexible model

Challenge 6: Backwards compatibility between releases

What Happens

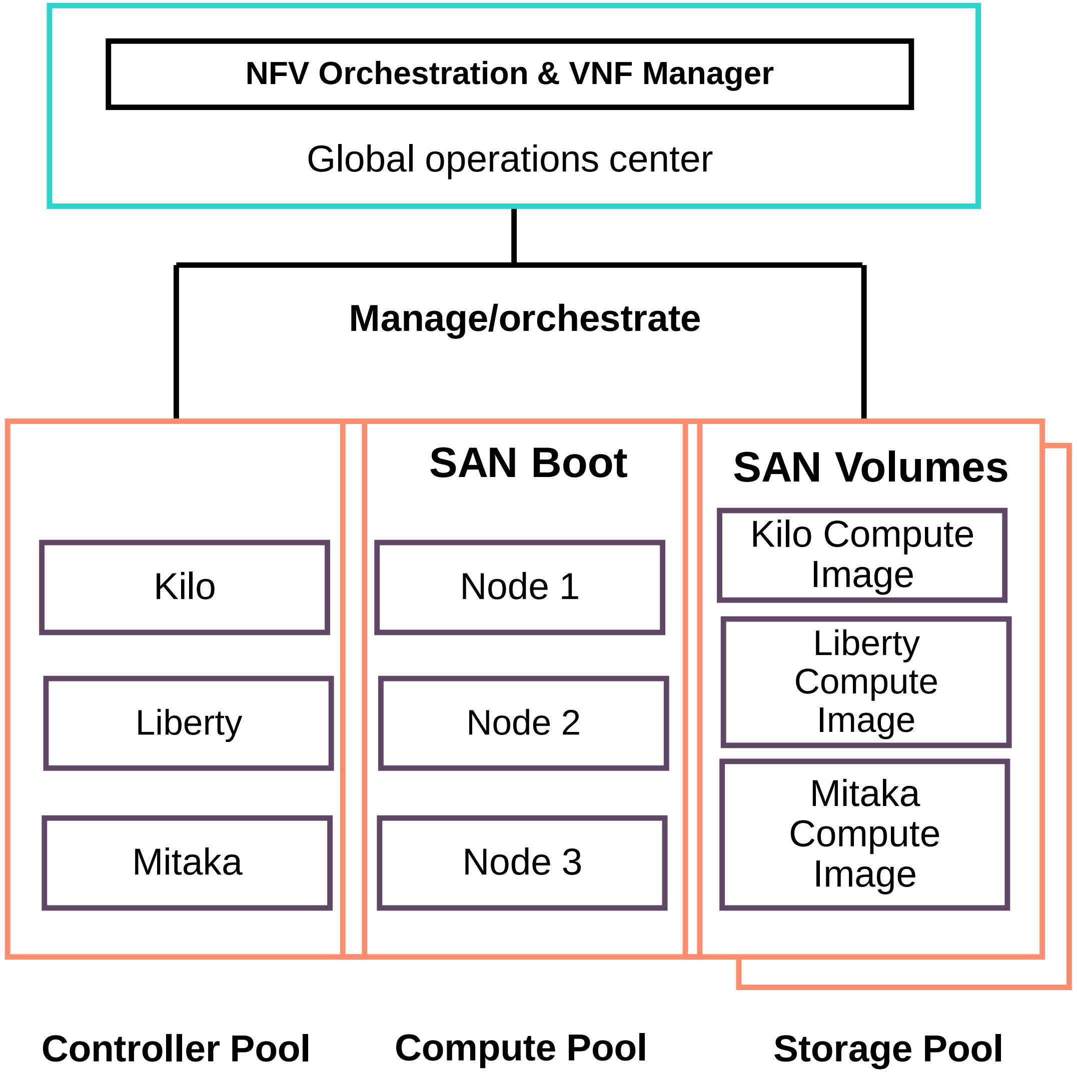
- Customer using a VNF built to work with a certain version of OpenStack
- Customer VNF upgrades are fairly limited in number; Stable product
- OpenStack releases upgrades on a bi-annual basis
- New OpenStack release deprecates APIs
- New OpenStack release fixes a critical security bug and/or adds an NFV enhancement

Why Is It Important

- Customer cannot upgrade to newer version since VNF incompatible with OpenStack version
- Customer systems exposed to security vulnerabilities
- Multiple engineering windows impact system upgrades for multiple customers
- Customer systems non-competitive due to key missing feature

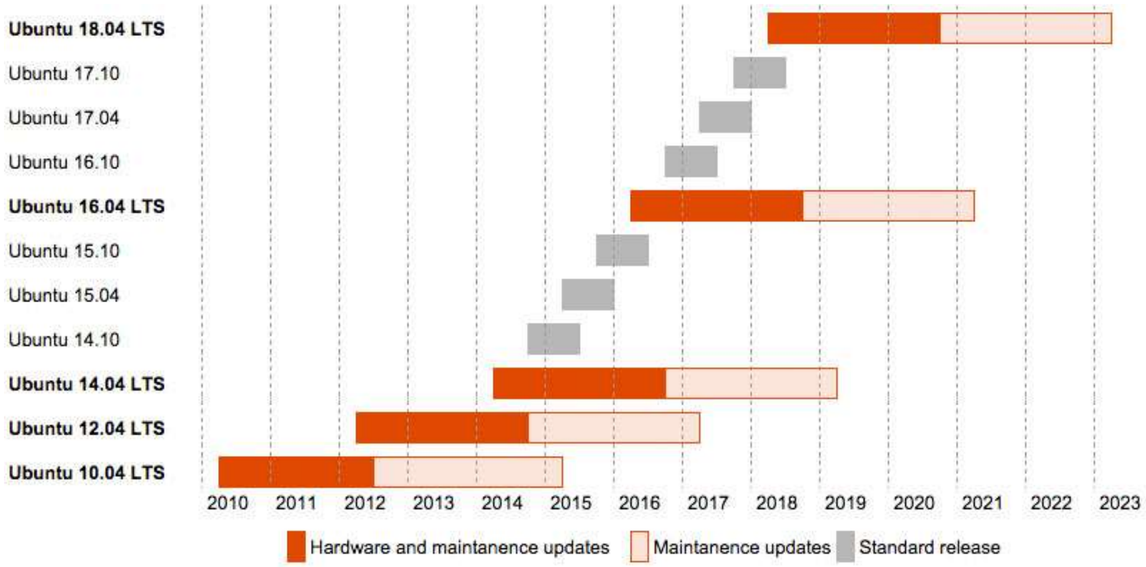
Short Term Solution 6: Backward Compatibility

- **Backport critical and domain specific features**
  - Community driven upstream efforts
  - Vendor supported product efforts
  - Customer driven solution efforts
- **Remote storage options**
  - Boot compute nodes from remote volumes
  - Multiple OpenStack versions as bootable volumes
  - VNFs available from bootable volumes or image repositories
- **Deprecation Countdown Timers**
  - Indication to developers on APIs up for deprecation
  - Pathway to move to newer, compatible APIs



Long Term Solution 6: Backward Compatibility

- **Safe Harbor Release**
  - Long Term Support for a “Safe Harbor Release”
  - Provide seamless upgrade and migration capability from one Safe harbor release to another
- **Cloud Portability Kits**
  - API layers to abstract OpenStack API layers
  - Consistent and long term support for Northbound APIs for Cloud Portability kits



Solutions Summary

Binding Virtual Network Interface Cards to the Virtual Network Function

Consistent Device Naming.
Nova boot with meta-data.
Virtual guest device role tagging.
Image Property PCI mappings

Now 2018

Service chain modification

Networking-sfc – port group mechanism
Advanced SFC capabilities in Neutron.

Now 2018

Start-up Storms (Or Stampedes)

Dynamic scaling of control instances.
High Availability solutions.
Services at the edge.

Now 2018

Securing OpenStack over the Internet

VPN outside of OpenStack control.
Federated local & remote shared services

Now 2018

Scalability of the controller(s)

Lightweight hypervisor manager.
Federated local & remote shared services.
Kolla & TriCircle.

Now 2018

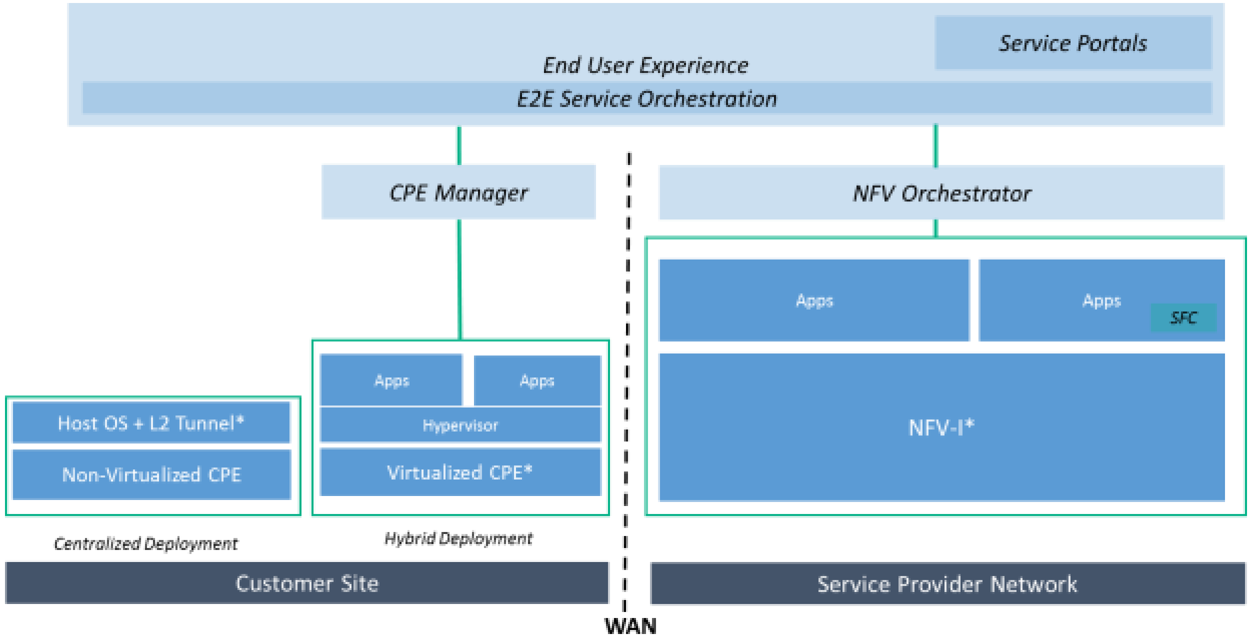
Backwards compatibility between releases

Backport critical features, remote storage.
Deprecation countdown timers
Safe harbour releases
Portability kits.

Now 2018

Conclusion & Call For Action

- There are several tractable & competing solutions to the 6 challenges for D-NFV at various stages of maturity.
- Making progress but not quick enough as Network Operators are launching NFV services now.
- It’s OK to have a limited number of competing solutions.
- Call For Action: Specifically we need help to prioritise:
  - Kolla & TriCircle development
  - Networking-sfc development
- Engage with OpenStack & OPNFV communities to make D-NFV challenges mainstream
- Need continued operator engagement – vendors can’t solve it alone.
- What challenges did we miss?



Hewlett Packard Enterprise

Hewlett Packard
Enterprise

intel®

BT

Thank you

Intel Legal Notices and Disclaimers

- Intel technologies' features and benefits depend on system configuration and may require enabled hardware, software or service activation. Learn more at intel.com, or from the OEM or retailer.
- No computer system can be absolutely secure.
- Tests document performance of components on a particular test, in specific systems. Differences in hardware, software, or configuration will affect actual performance. Consult other sources of information to evaluate performance as you consider your purchase. For more complete information about performance and benchmark results, visit **http://www.intel.com/performance**.
- Intel, the Intel logo and others are trademarks of Intel Corporation in the U.S. and/or other countries. *Other names and brands may be claimed as the property of others.
-