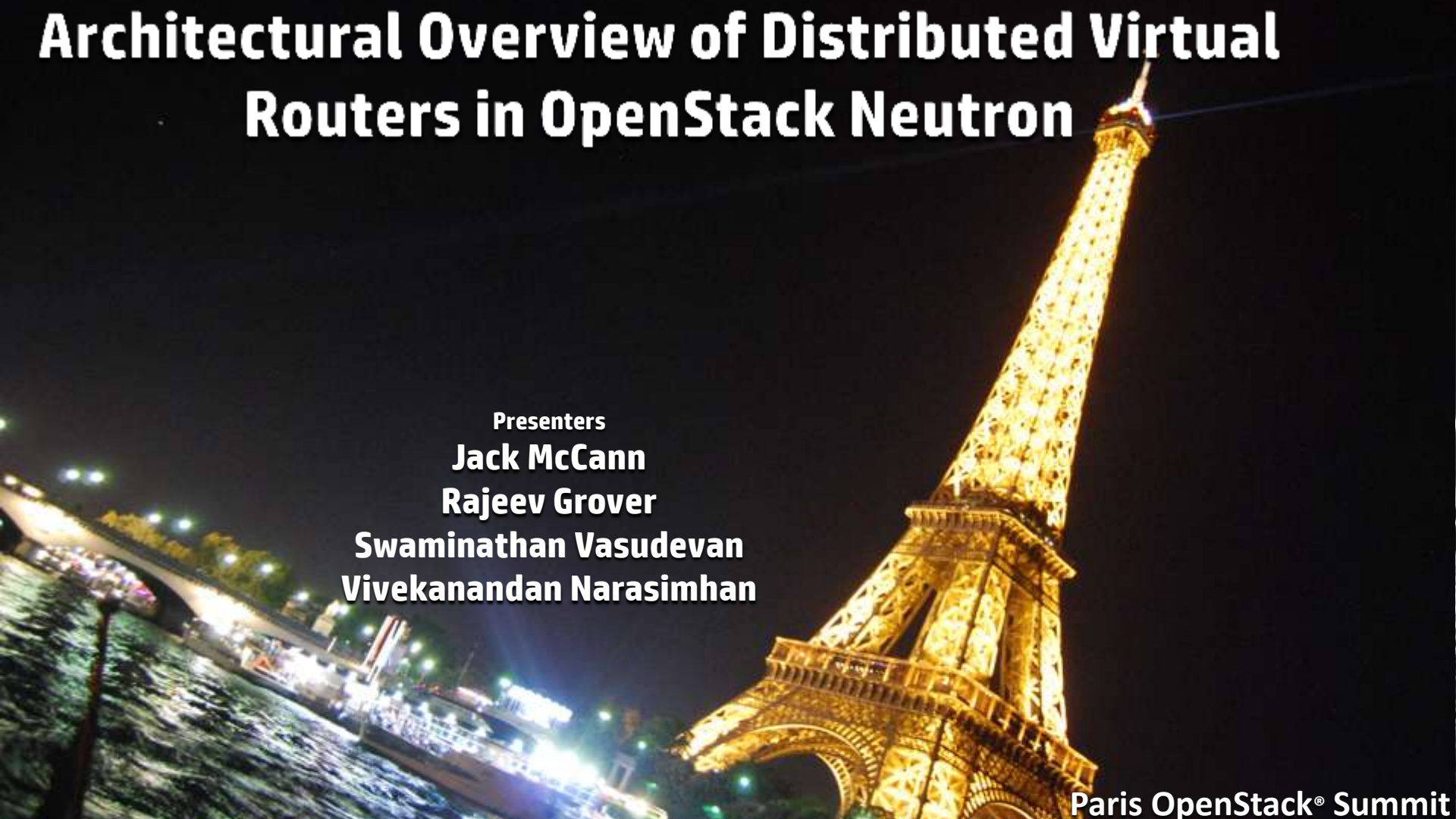
Architectural Overview of Distributed Virtual Routers in OpenStack Neutron



Presenters

Jack McCann

Rajeev Grover

Swaminathan Vasudevan

Vivekanandan Narasimhan

Paris OpenStack® Summit

Agenda

- Introduction
- High level architecture and DVR Configuration
- North-South Routing
- East-West Routing
- OVS Rules
- Scheduling
- Services
- API Changes and DB Extensions
- Future Plans for DVR

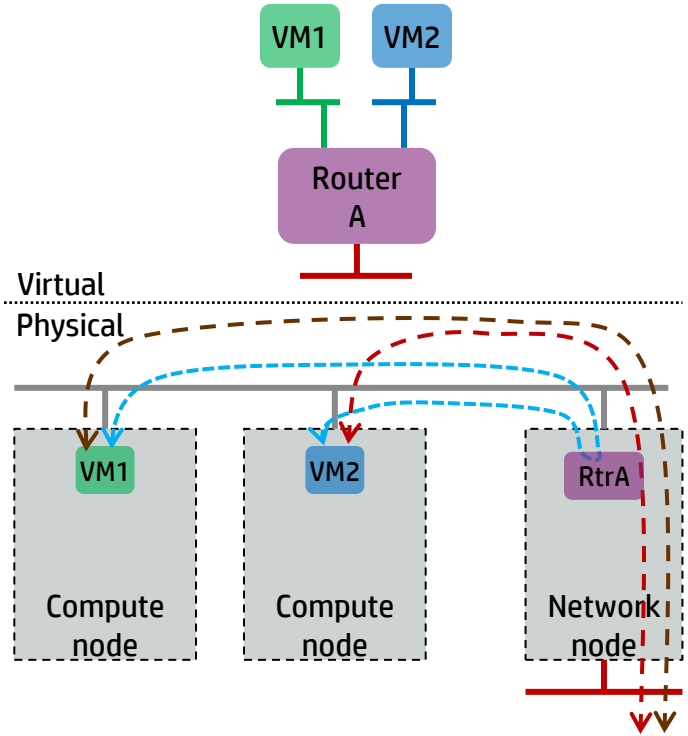
Legacy Routing in Neutron

Network node provides:

- IP forwarding
  - **Inter-subnet** (east-west) traffic between VMs
  - **Floating IP** (north-south) traffic between external and VM
  - **Default SNAT** (north-south) traffic from VM to external
- Metadata Agent
  - access to Nova metadata service

Issues:

- Performance bottleneck
- Scalability limitations
- Single Point of Failure



Distributed Routing in Neutron

Compute node provides:

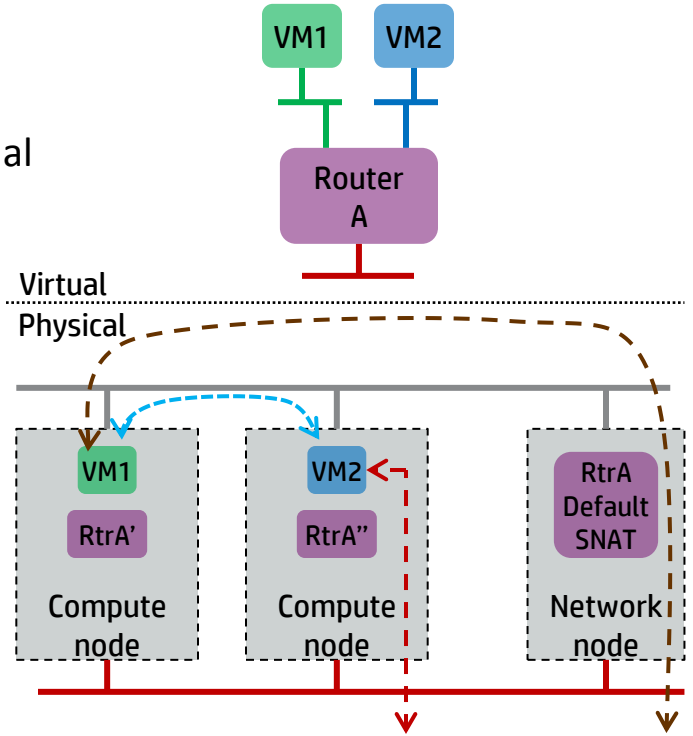
- IP forwarding for local VMs
  - **Inter-subnet** (east-west) traffic between VMs
  - **Floating IP** (north-south) traffic between external and VM
- Metadata Agent for local VMs
  - access to Nova metadata service

Advantages:

- Bypass network node improves performance
- Scales with size of compute farm
- Limited failure domain (per compute node)

Limitations:

- **Default SNAT** function is still centralized

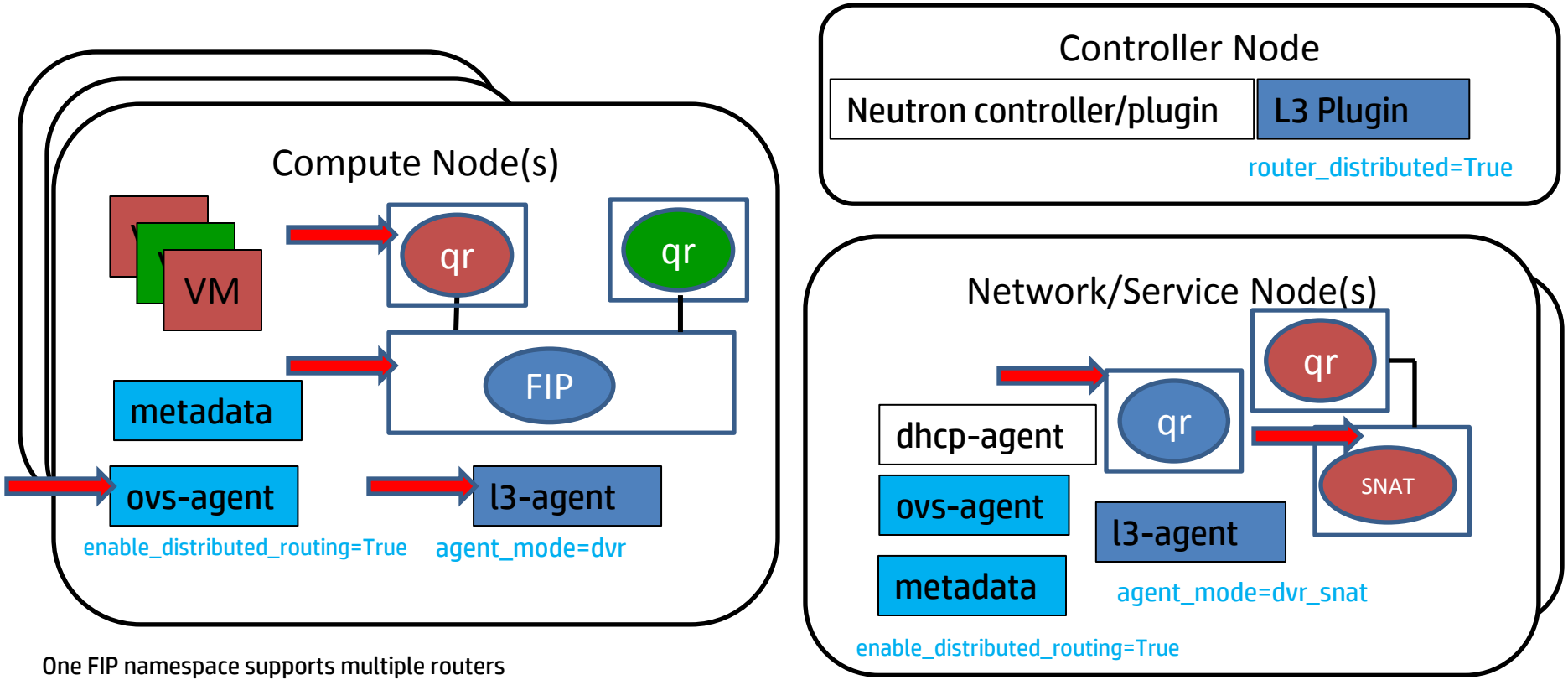


High level requirements for DVR

- Help close the parity gap with Nova (multi-host)
- Provider feature
  - Tenants should not have to know or care
- Configurable on a per-router basis (centralized or distributed)
  - Default router type set by a global config knob
- Can be deployed into existing environments
- Centralized routers and Distributed routers can coexist in same cloud
- Ability to migrate a router from Centralized to Distributed
- Minimize overhead use of public IP addresses
- Leverage existing code base

DVR High Level Architecture and Configuration

DVR High level Architecture



- One FIP namespace supports multiple routers
- Legacy OpenStack Routers (non-distributed) can exist on the network/service node
- Enhanced l3-agent works in different modes “dvr_snat”, “legacy” and “dvr”.

Configuring DVR

DVR Configuration Overview

Plugin Configuration (Default Router Option)

```
"neutron.conf"
router_distributed=True
```

L3 Agent Configuration and Modes

```
"l3_agent.ini"
agent_mode=dvr_snat ( Network/Service Node)
agent_mode=dvr (Compute Node only)
agent_mode=legacy ( Network/Service Node)

router_delete_namespaces=True ( Enable namespace
cleanup)(Optional)
```

DVR Configuration Overview

L2/OVS Agent Configuration

"ml2_conf.ini"

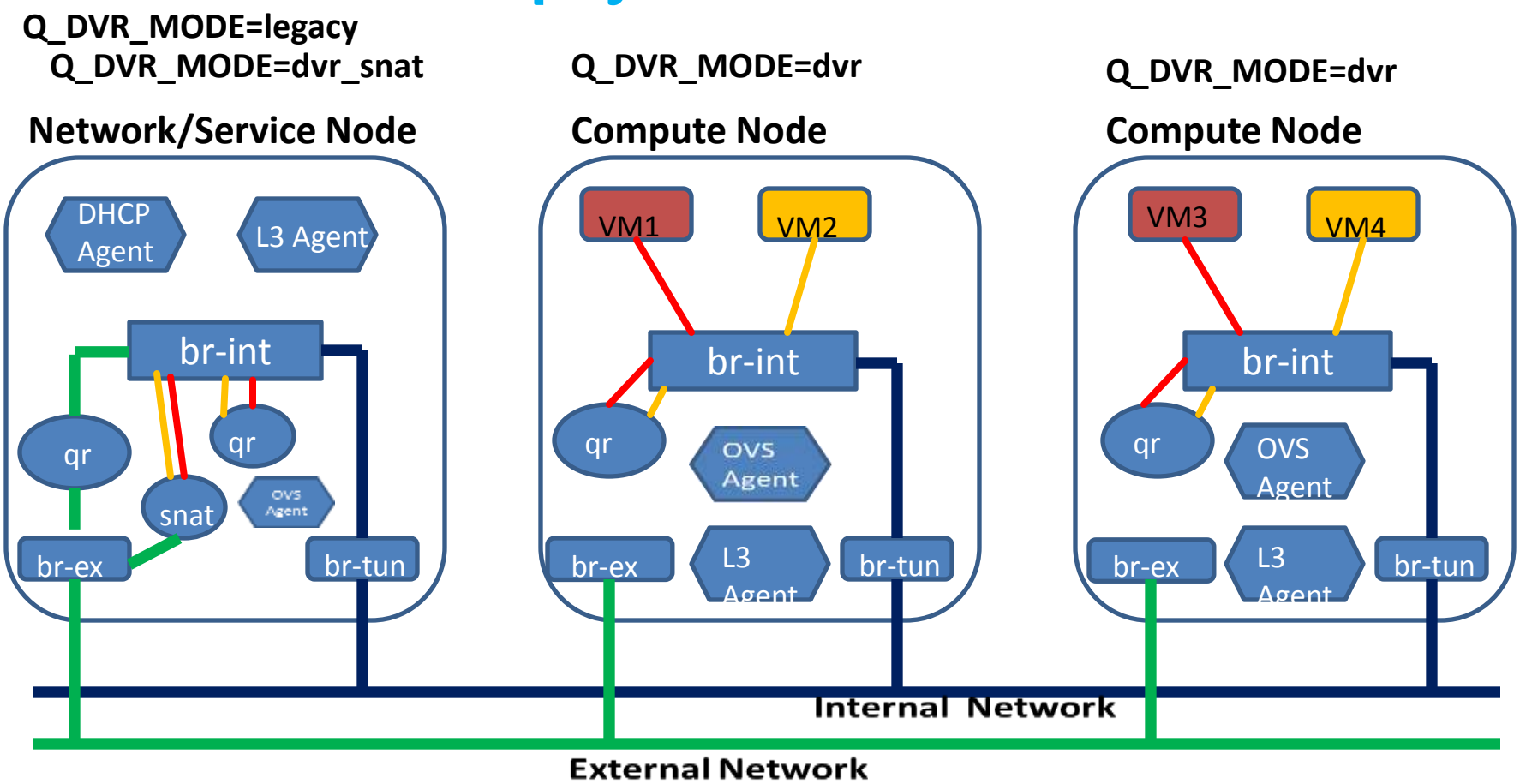
```
enable_distributed_routing = True

enable_tunneling = True
local_ip = <Local Data Network IP (or) TEP IP>

tunnel_types = vxlan
l2_population=True
```

DevStack Configuration

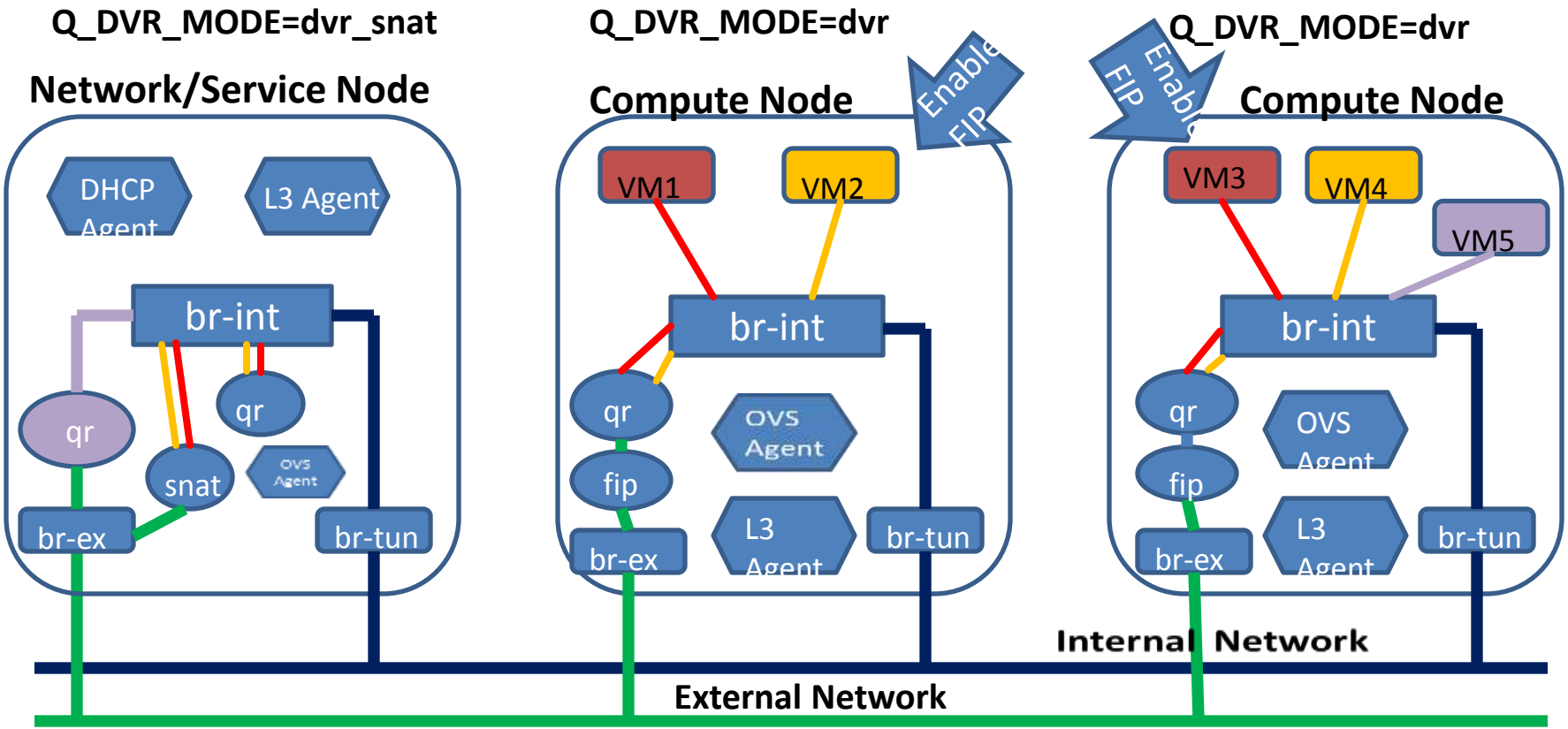
"local.conf"

```
Q_DVR_MODE=dvr_snat/dvr/legacy
```

Legacy Deployment
DVR Deployment without FIP

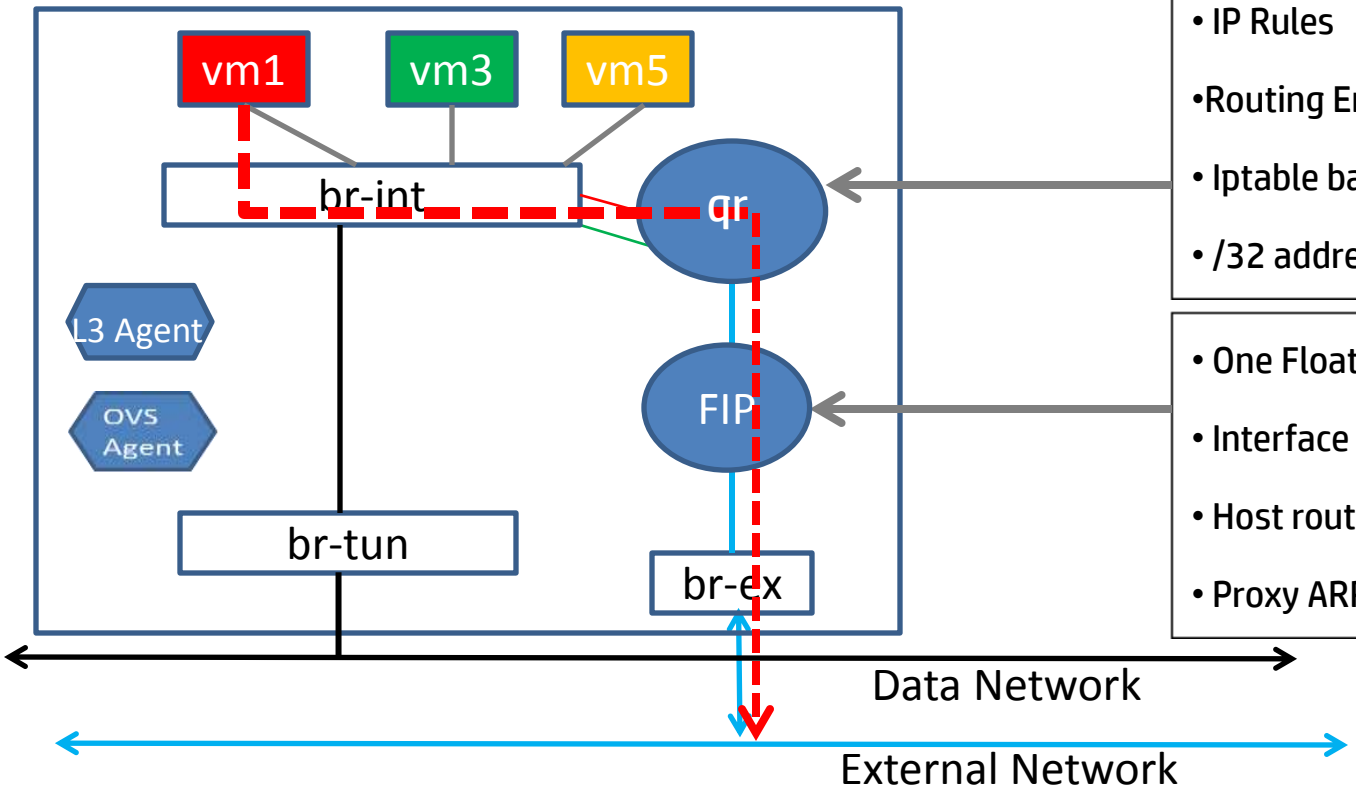


DVR Deployment with FIP



North-South Routing

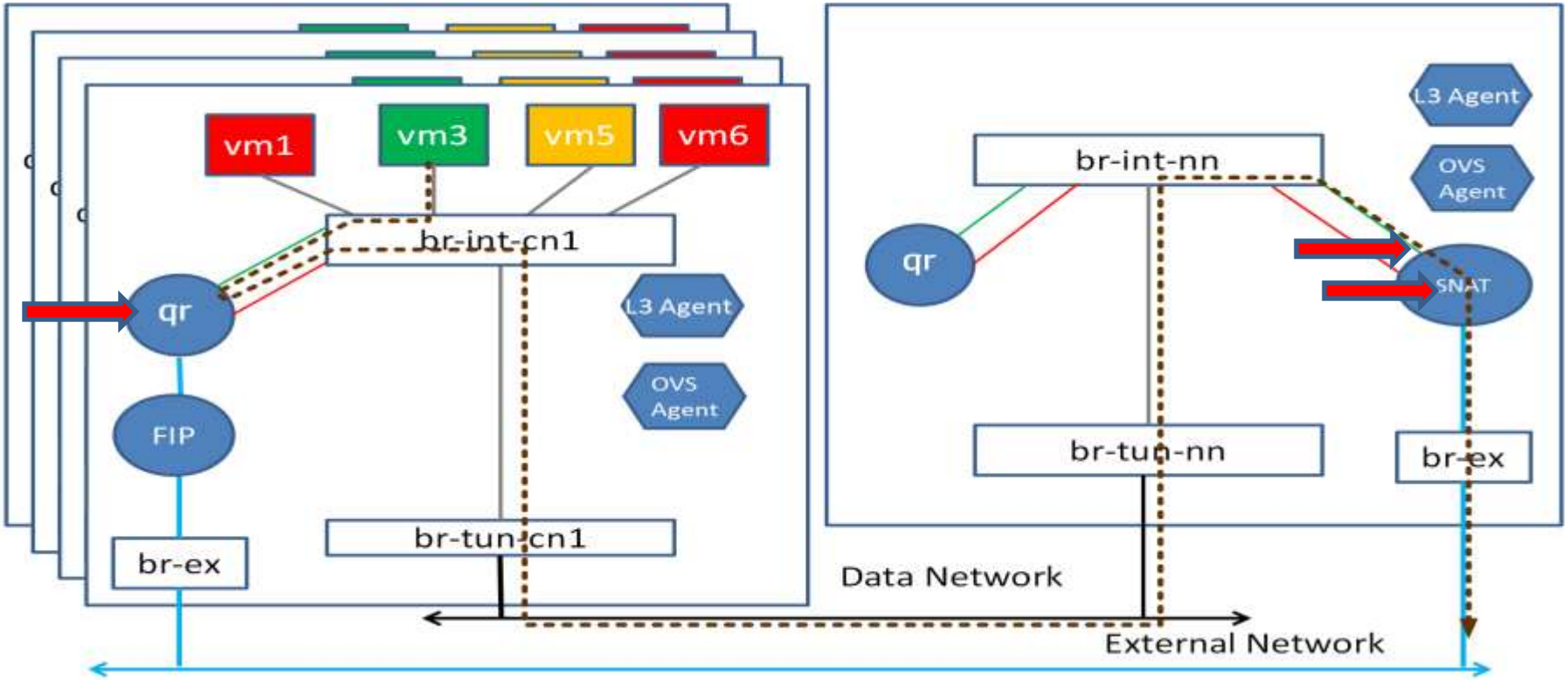
Overview of North-South Routing



- IP Rules
- Routing Entries
- Iptable based NATing rules
- /32 addresses for the FIP

- One Floating IP Namespace per CN
- Interface on external network
- Host routes to qr for FIP
- Proxy ARPs enabled for FIP

Default SNAT Traffic flow



East-West Routing

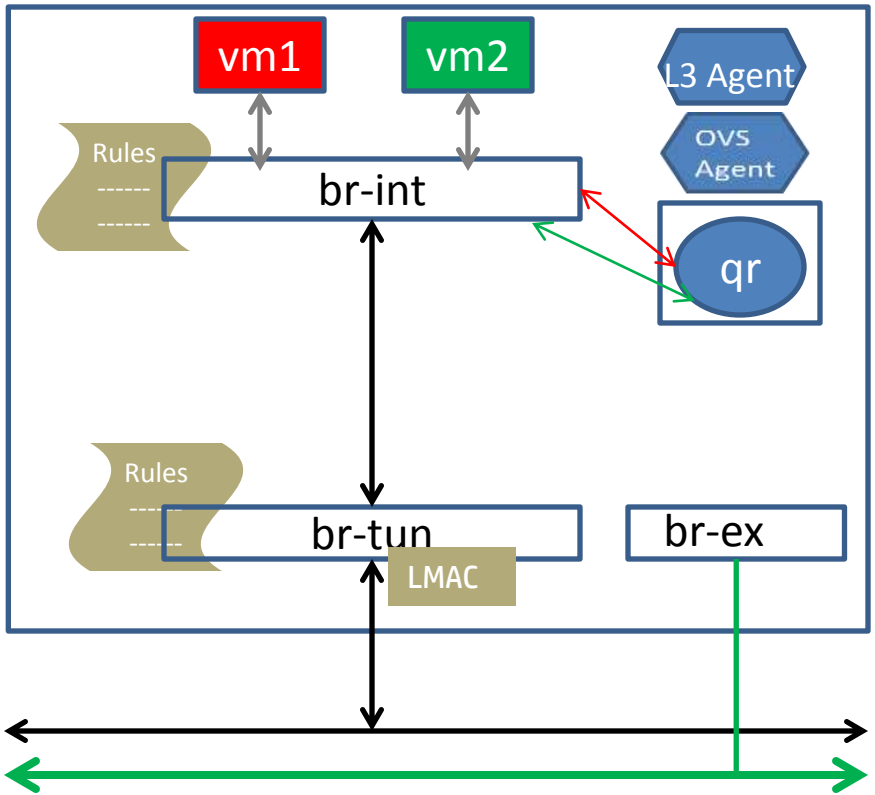
Compute Node entities for enabling the DVR East West Routing

East-West routing is accomplished through a combination of the following in each CN:

•**qr** : a namespace that forwards traffic among routed subnets using route table entries.

•**LMAC**: a MAC address unique to each CN for use as source address in forwarding routed frames from one CN to another.

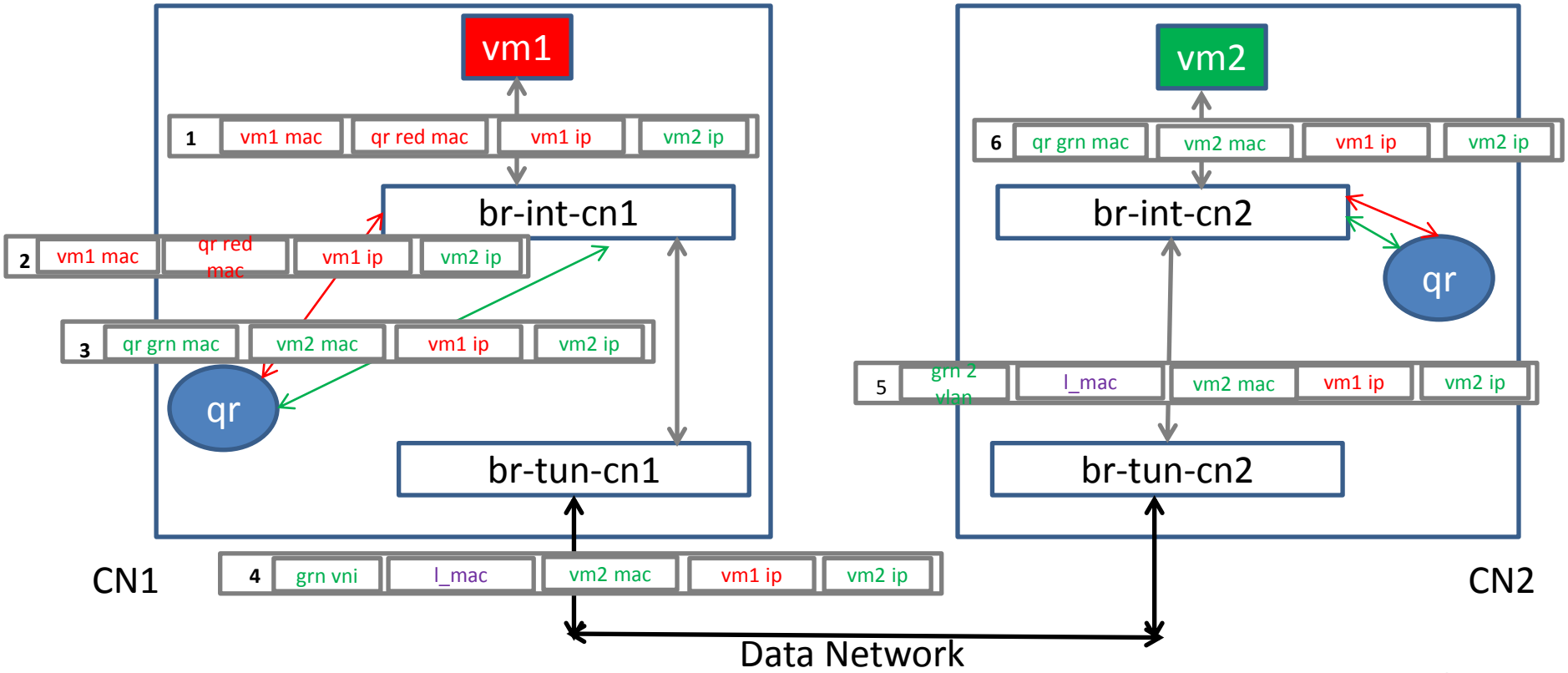
•A set of rules in the OVS bridges that prevent frames with src=gateway MAC from egressed tunnel bridge . These rules also cause peer bridges to substitute/restore gateway MACs with LMACs while routed frames traverse through br-tuns and underlay network.



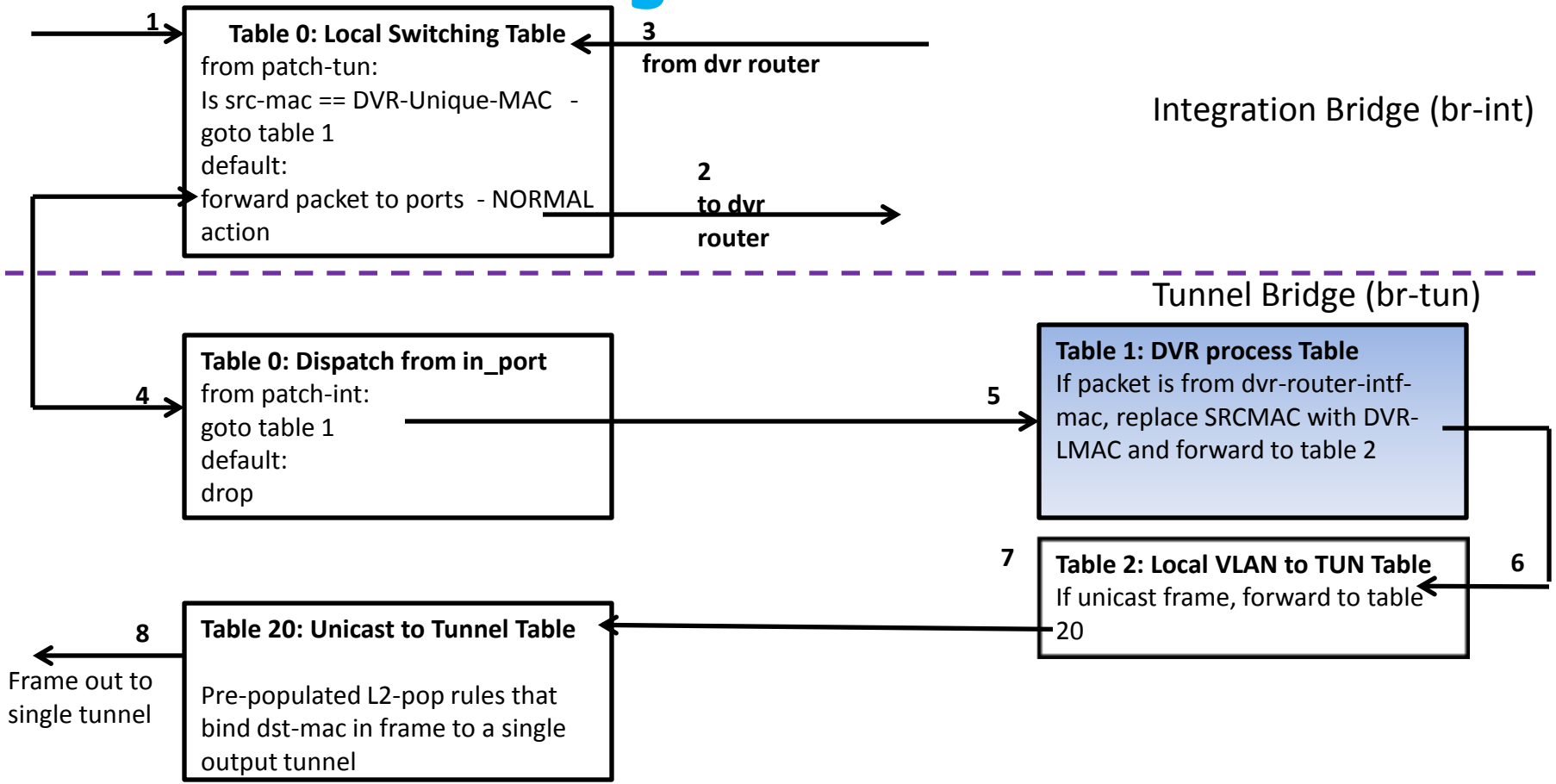
Flow of DVR routed packets (example)

Tenant has two VMs - vm1 that is in RED Net & vm2 that is in GREEN Net
CN1 and CN2 are two compute nodes. qr represents a dvr owned by Tenant 1

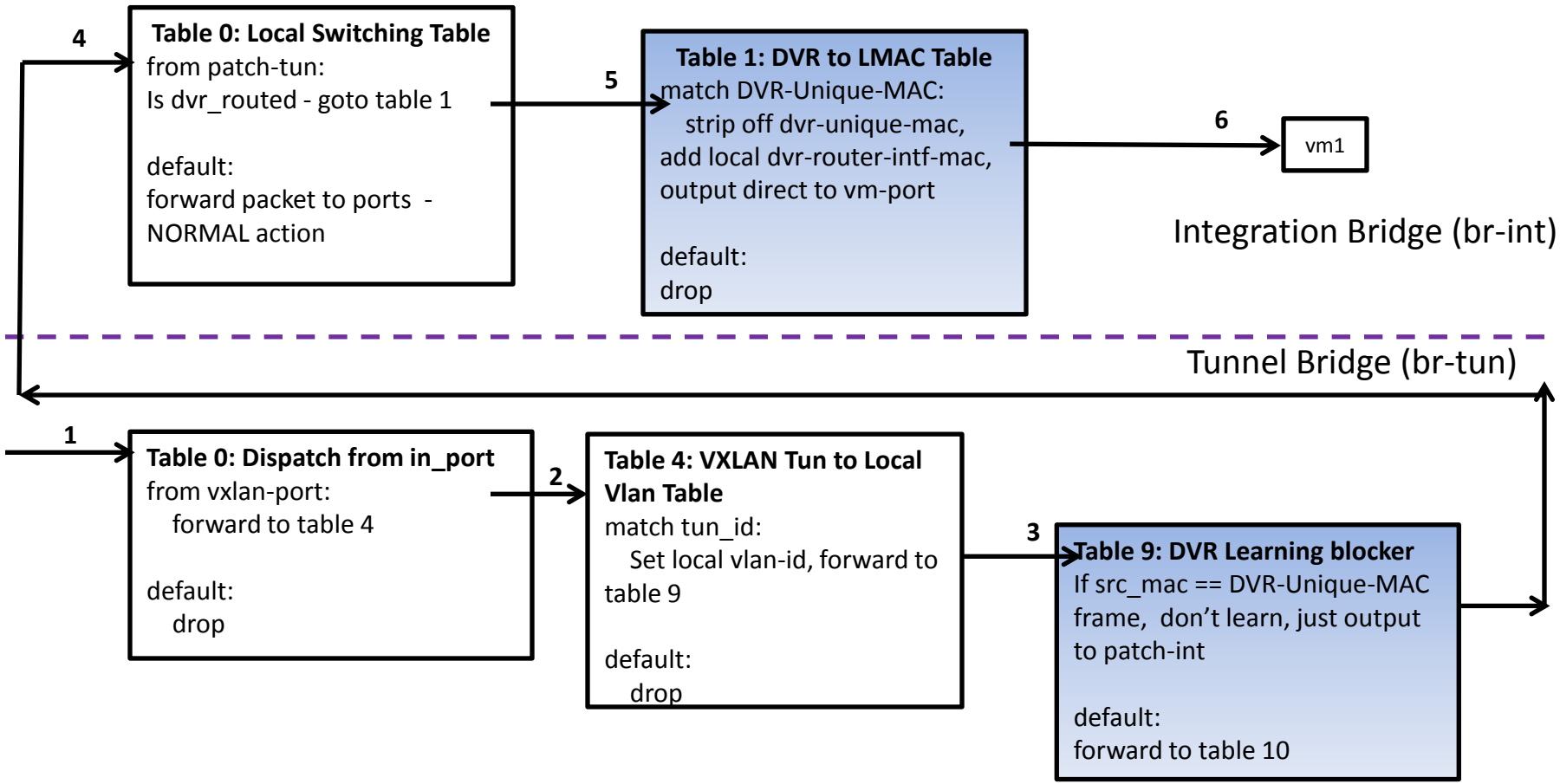
PING REQUEST from vm1 to vm2



Egress to Cloud

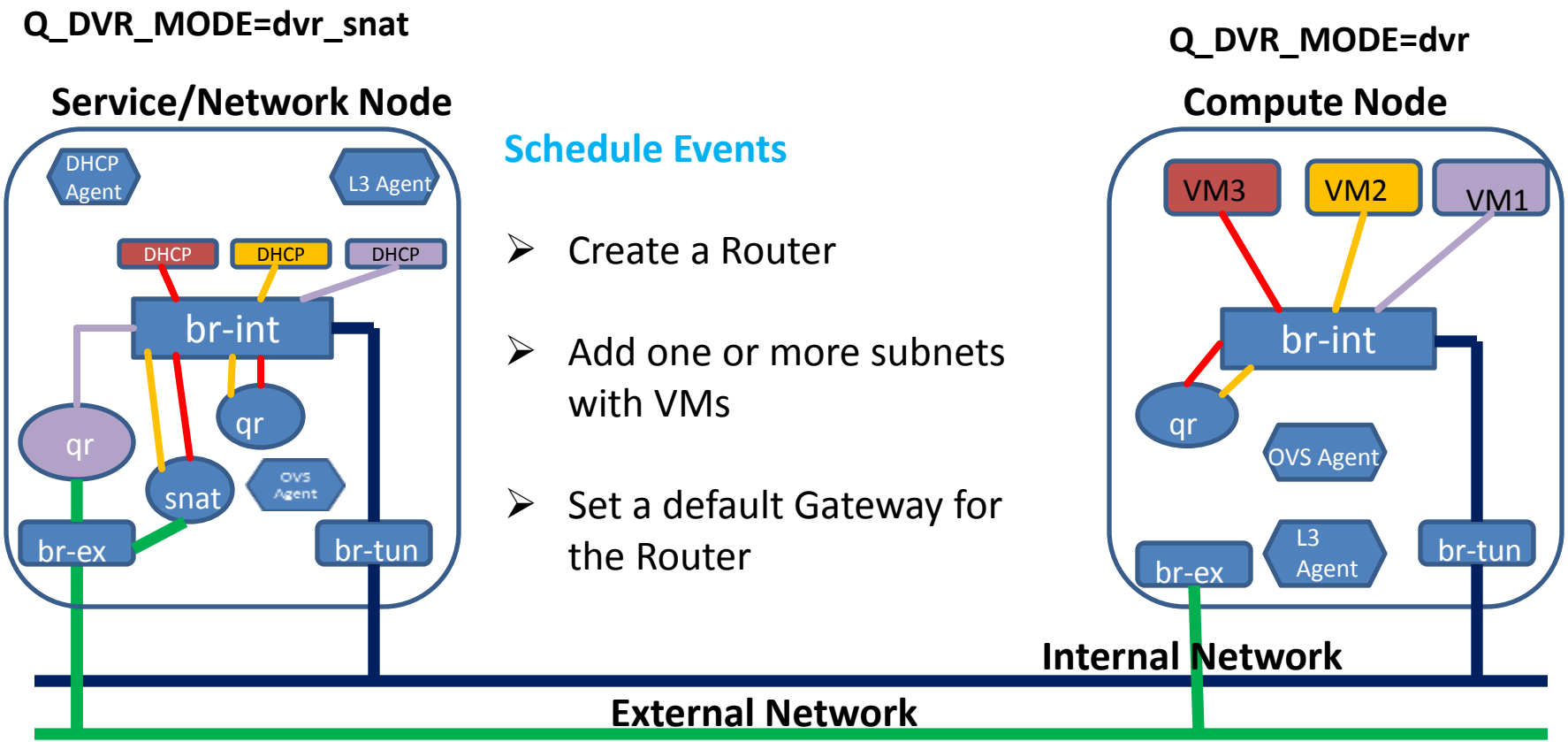


Ingress from Cloud

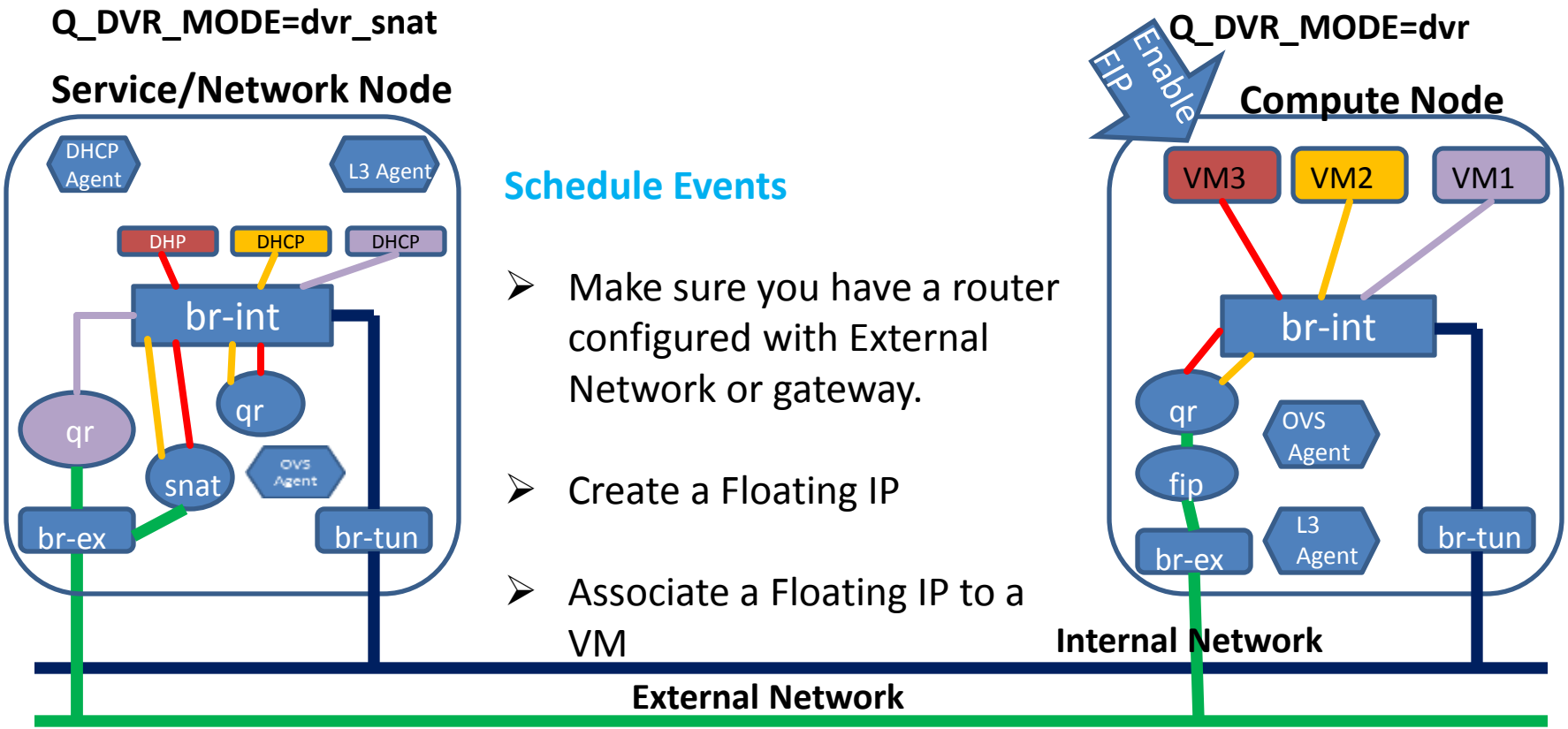


Scheduling

Scheduling a “qr/snat” in Service/Compute Node



Scheduling a “fip” in Compute Node



Un-Scheduling Routers and FIP

Un Schedule Events

FIP Namespace cleanup

- Last VM holding the FIP determines the FIP namespace cleanup
- Delete a FloatingIP
- Disassociate a Floating IP

Router Namespace cleanup

- When no more DVR related ports (Includes Compute, VIP, DHCP etc.,) are serviced by the routed subnet, the router namespaces are cleaned up.
- Both l2 agent and l3 agent takes part in the clean up.

SNAT Namespace cleanup

- When a gateway is removed from a router the snat namespace will be cleaned up.

NOTE: Namespace cleanup should be enabled “router_delete_namespaces=True” in l3_agent.ini

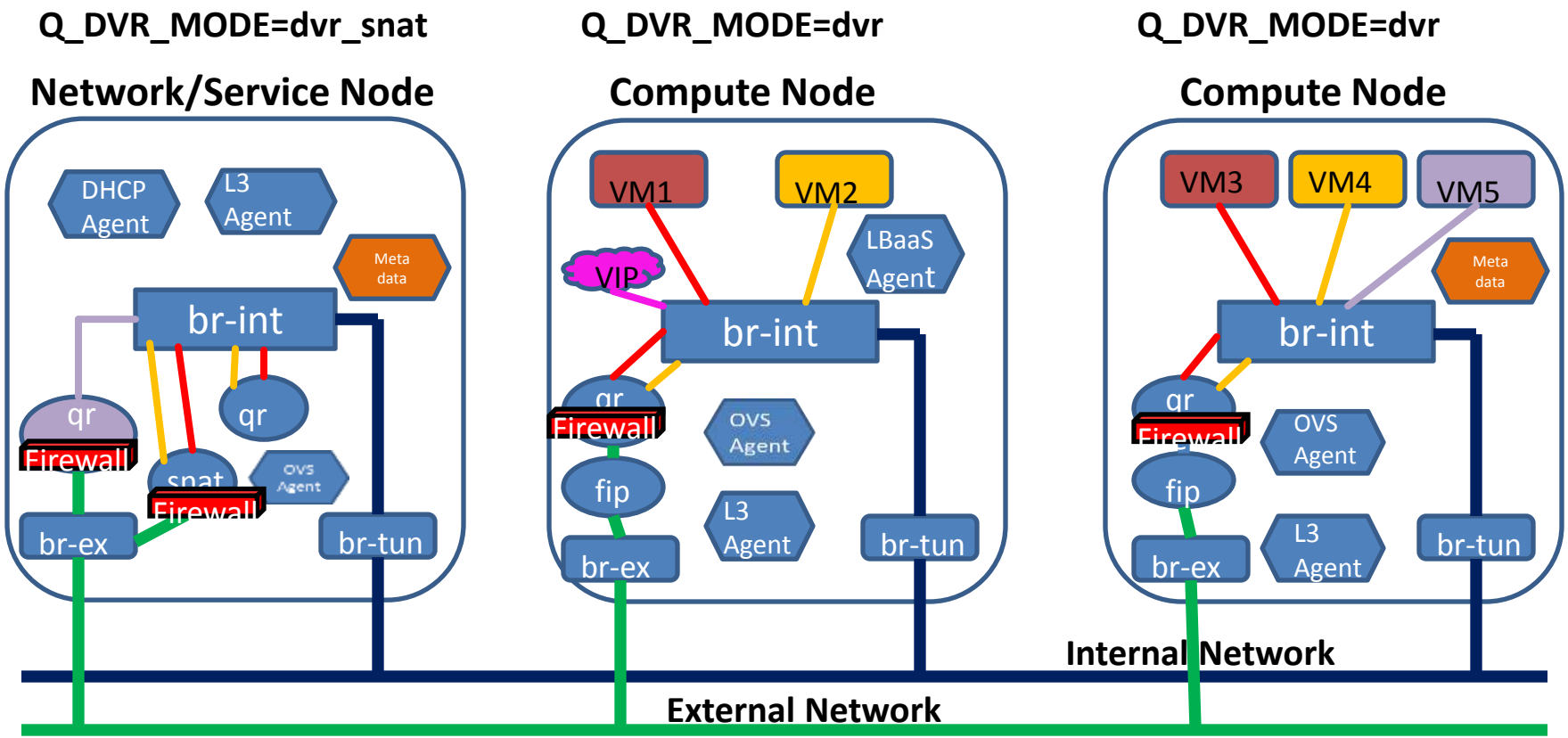
Services

Services support with DVR

Services Support in DVR as of Juno

- LBaaS
- FWaaS
  - North-South only (External Traffic)
  - East-West (Internal Traffic)
- MetadataService
- VPNaaS is still supported with Centralized Routers.
- VPNaaS support for DVR is Work in Progress.

Services deployment with DVR



API Changes and DB Extensions

API Extensions

Adds ‘distributed’ attribute to the ‘router’ object

Router Create:

```
neutron router-create --distributed=True/False
```

Router Update:

```
neutron router-update --distributed=True/False
```

Can be set by admin user through the above API
Global default is set as “router_distributed” in neutron.conf
The attribute is only visible to admin tenant in GET.
NOTE: Migrating or Converting a Legacy Router to Distributed is Work in Progress.

DB Changes for DVR

router_extra_attributes

router_id	string uuid
distributed	boolean

dvr_host_macs

host	string 255
mac_address	string 32

csnat_l3_agent_bindings

router_id	string uuid
l3_agent_id	string uuid
host_id	string
csnat_gw_port_id	string uuid

ml2_dvr_port_bindings

port_id	string uuid
host	string
router_id	string uuid
vif_type	string
vif_details	string
vnic_type	string
profile	string
cap_port_filter	boolean
driver	string
segment	string
status	string

Future Plans

Future Plans for DVR

- VPNaaS support for DVR
- Full migration support for DVR routers.
- HA for Service Node
- IPv6 Support
- VLAN Support
- L3 Agent Refactor
- Distributed DHCP
- Performance tuning.
- Distributed SNAT

Questions?

THANK YOU