

DVR North South and East West

Changes from the IceHouse Summit Session Discussion:

NOTE: In the ICEHOUSE Summit Session we proposed that we would be providing a DVR solution only for the East-West Communication and this would be based on a DVR dataplane kernel Module Driver.

<https://docs.google.com/document/d/1kMUO1-y4yATQNBIKdjDu7OwFYt5cfzIzb5NKQjhb08/edit#heading=h.ttq3kyub25tn>

Action: Based on the feedback that we received from the ICEHOUSE Summit regarding support for North-South and also concerns regarding the use of the Kernel Module Driver, we moved away from our original kernel module design and had moved to a flexible model where we use the existing namespace routers operating in different mode (L3 or DVR).

Purpose:

To solve the scalability problem in the OpenStack Neutron Deployment and to distribute the Network Node load to other Compute Nodes.

Distributed Virtual Router solves both the problems by providing a solution that would fit into the existing model.

This document will discuss the design considerations for the Distributed Virtual Router that includes the East-West and North South communication.

The plan for North-South would be to provide Floating IP Support at the Compute Node for the required VMs.

Also provide a Default Centralized SNAT for other VMs that does not require a Floating IP. There are some use cases where the packets may still need to be forwarded to a centralized node with SNAT capability where singleton services such as VPN can be provided.

Limitations: For now, based on the current neutron design, the only service that will not be targeted for distributing is the "DHCP". So the current plan is to run the DHCP in a specific service node 'aka' legacy Network Node.

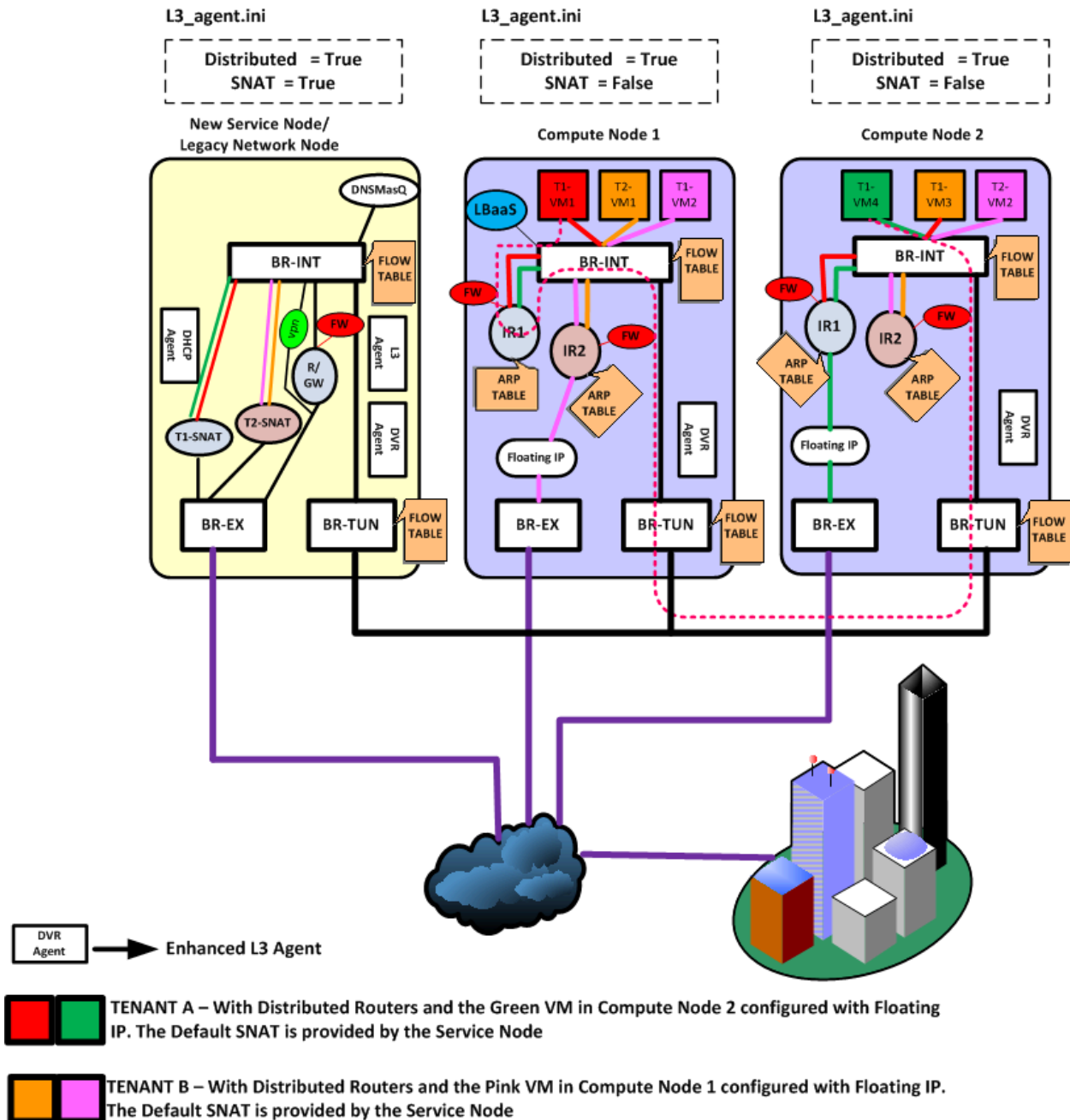
VPN is an exception and cannot be distributed. So this service will still be centralized in the service node or the legacy network node.

Target: The first release is targeted towards "OVS" based solution.

Use Cases:

Use Case 1: East-West:

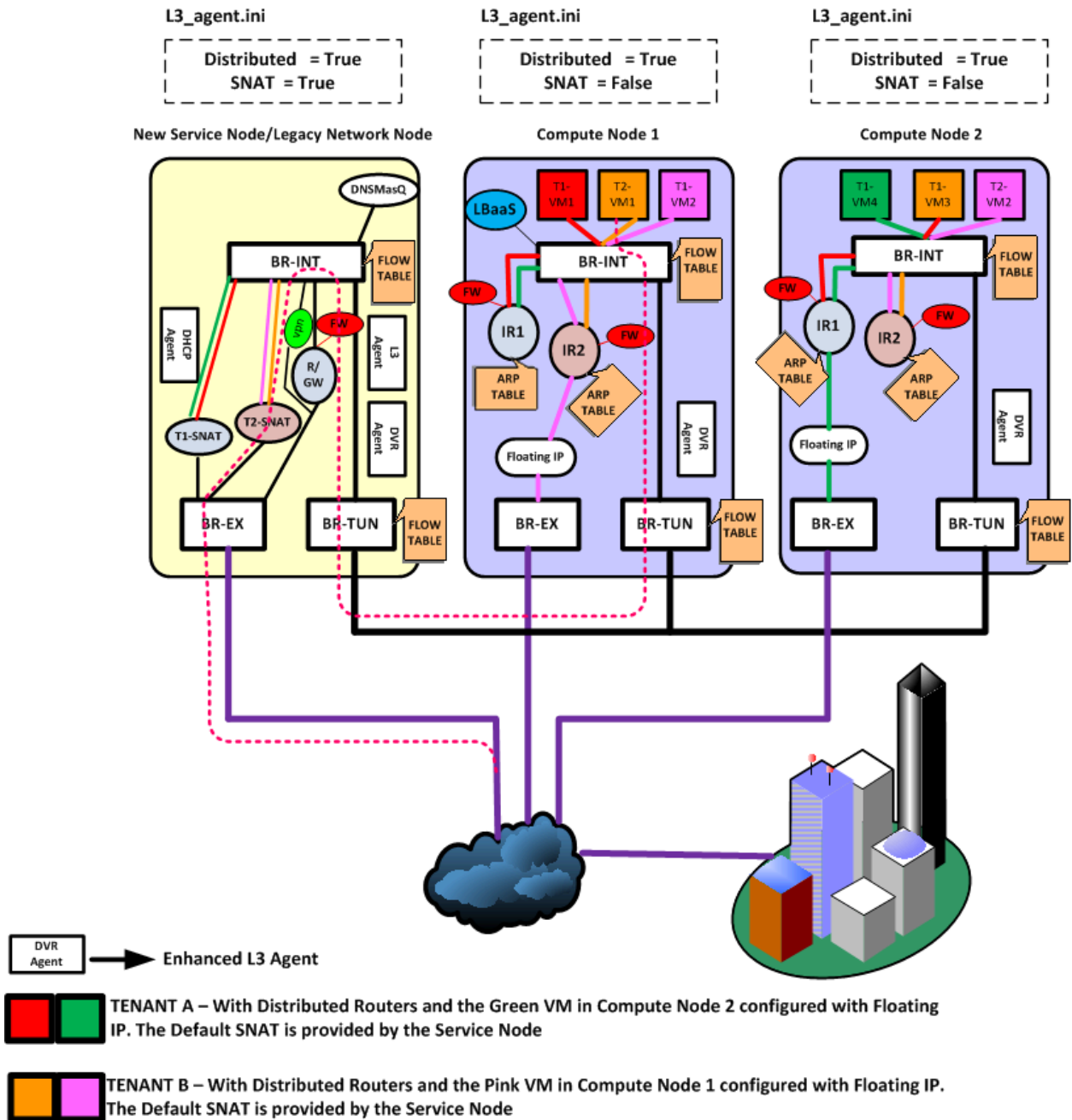
This would allow the intra-tenant traffic to flow between the compute Nodes without getting to the Network Node.



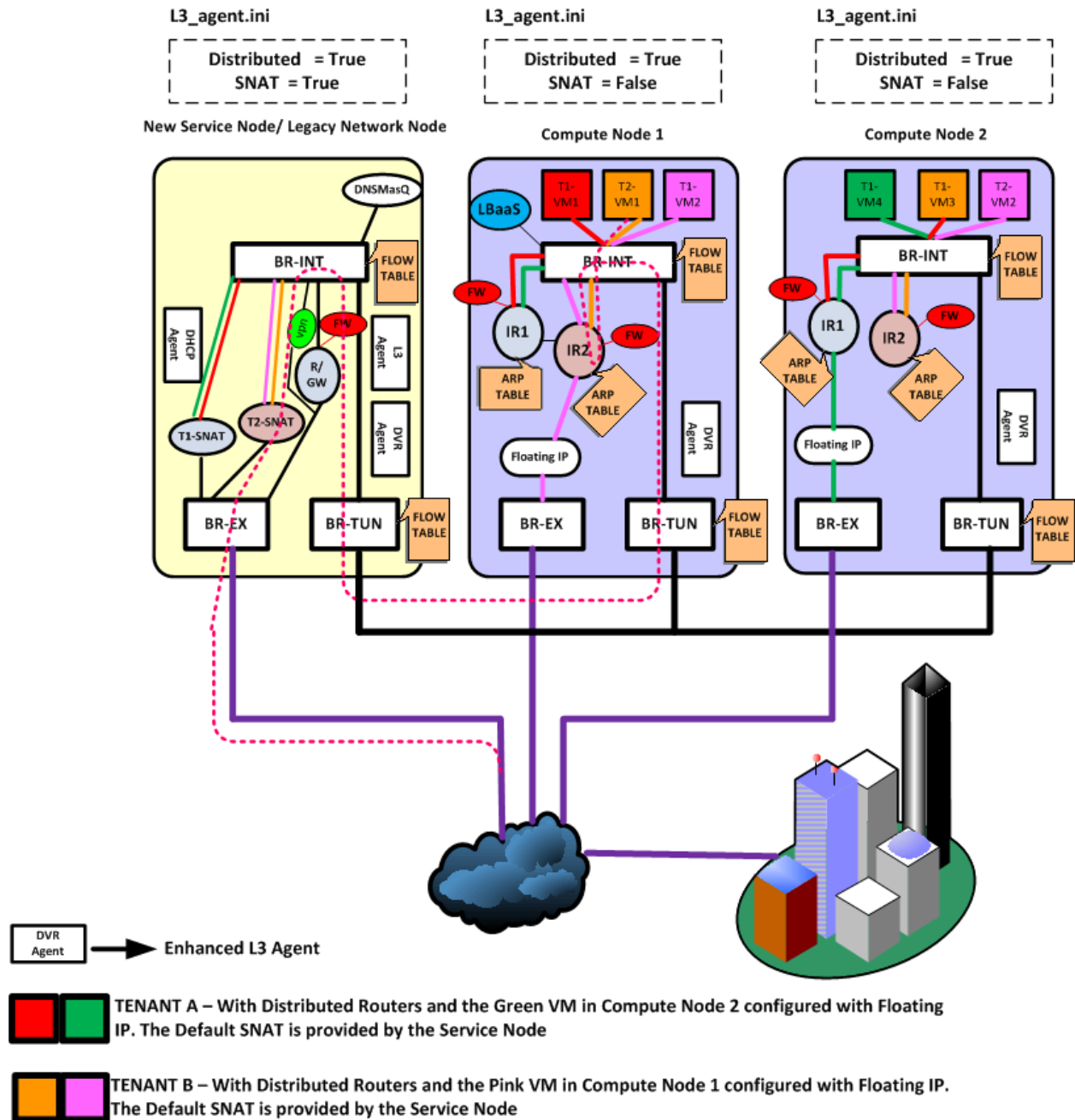
Use Case 2: East-West with Centralized default SNAT:

Cloud admins don't want to distribute the SNAT service to all the compute nodes, since it

consumes public IP Address and also there is requirement that certain services will be centralized that require SNAT.



The picture above shows the packet flow from External Public Network to a VM in an internal private Network through the Default SNAT.



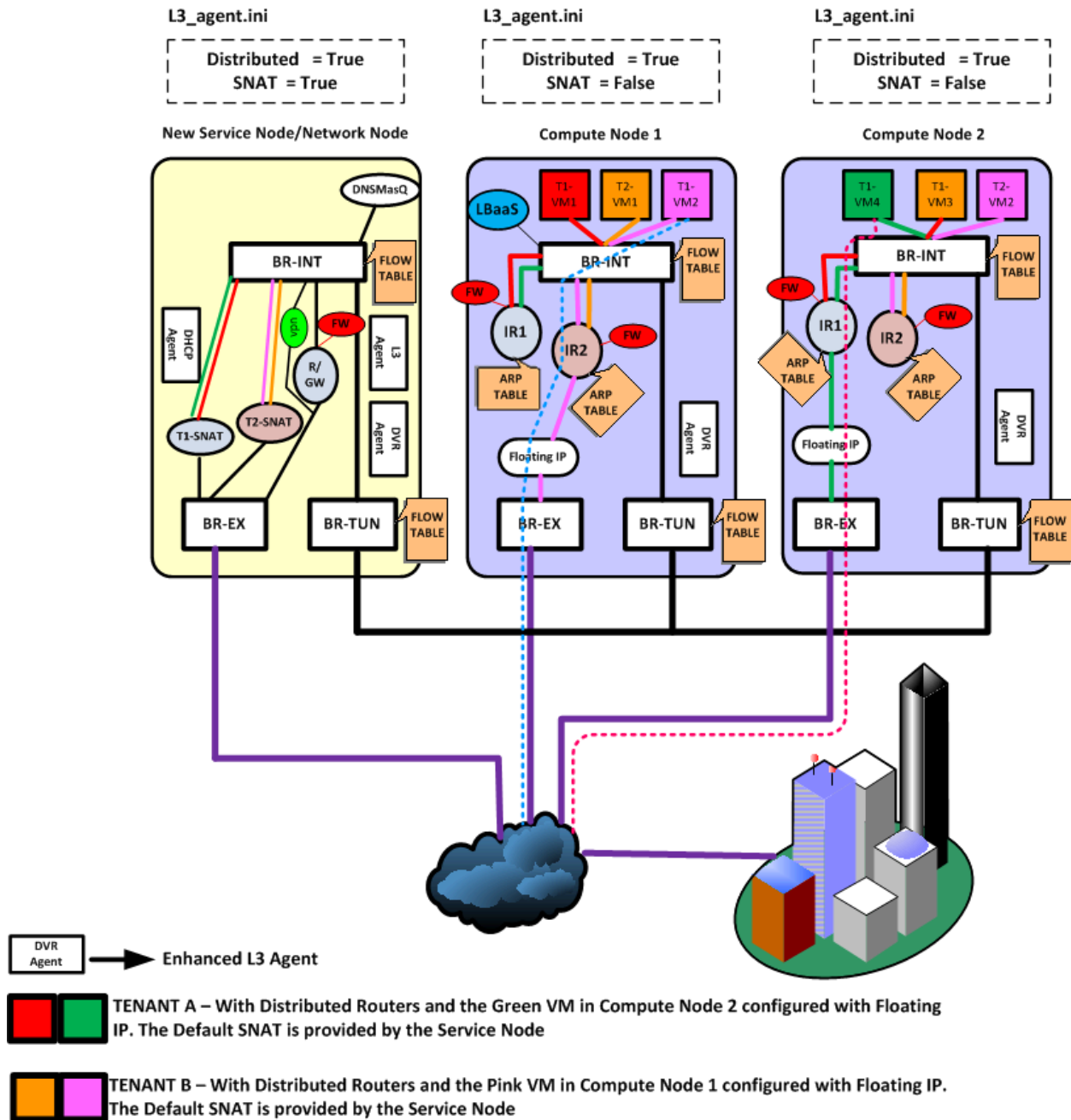
The picture above shows the packet flow from an internal VM in a private network to an External Network through the Default SNAT.

Use Case 3: North-South (Floating IP from the Compute Node)

Apart from the Default Centralized SNAT and East-West, Cloud admins/tenants want their VM traffic to hit the external network without going through the Network Node, since that would ease the Network Node from Load. Also this would facilitate in the future having service VMs

providing services that can be accessed from the external network.

In order to achieve this, the Floating IP allocation and translation will be done at the Compute Node rather than the Network Node.



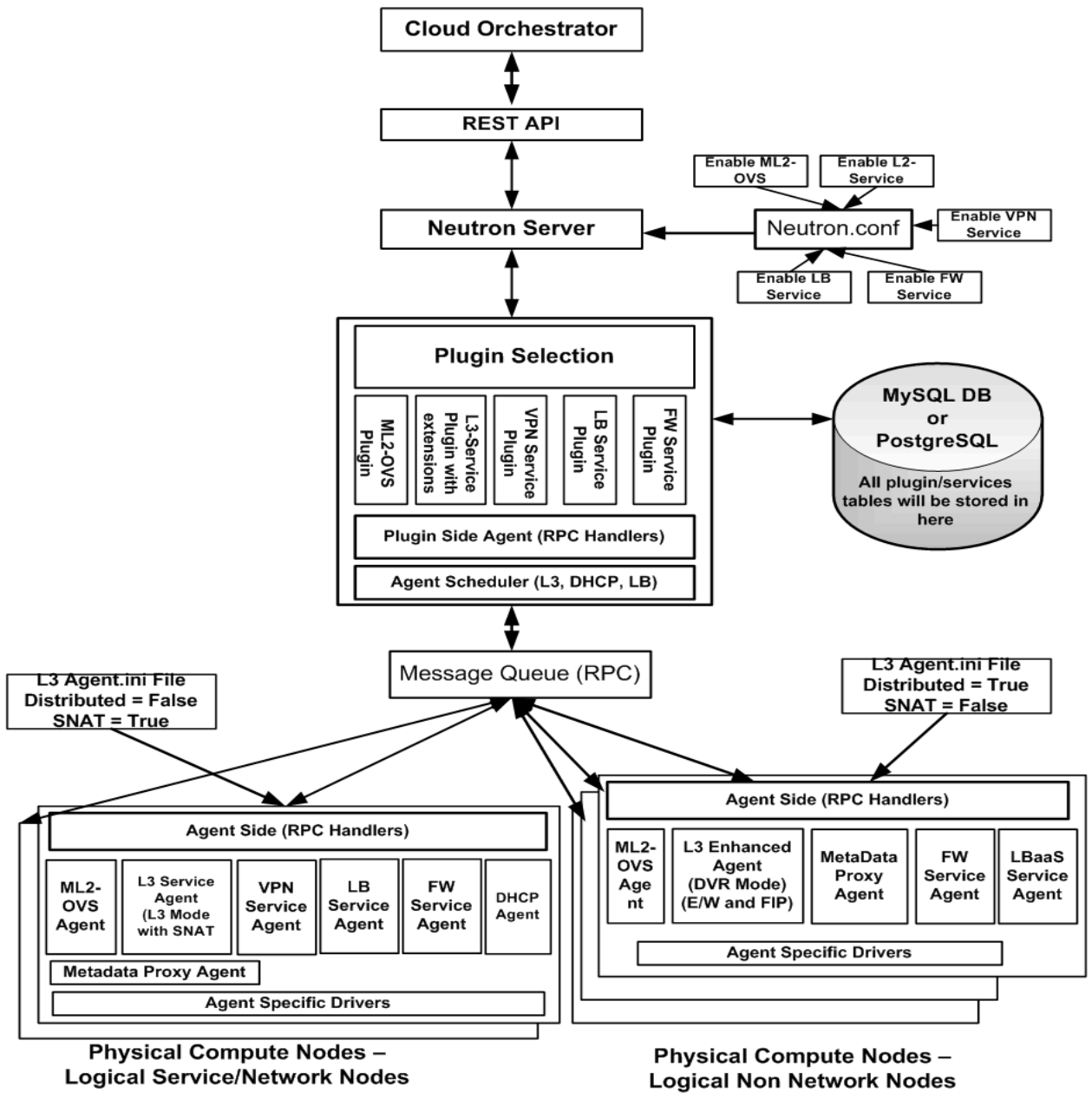
Definitions:

IR - Internal L3 Router

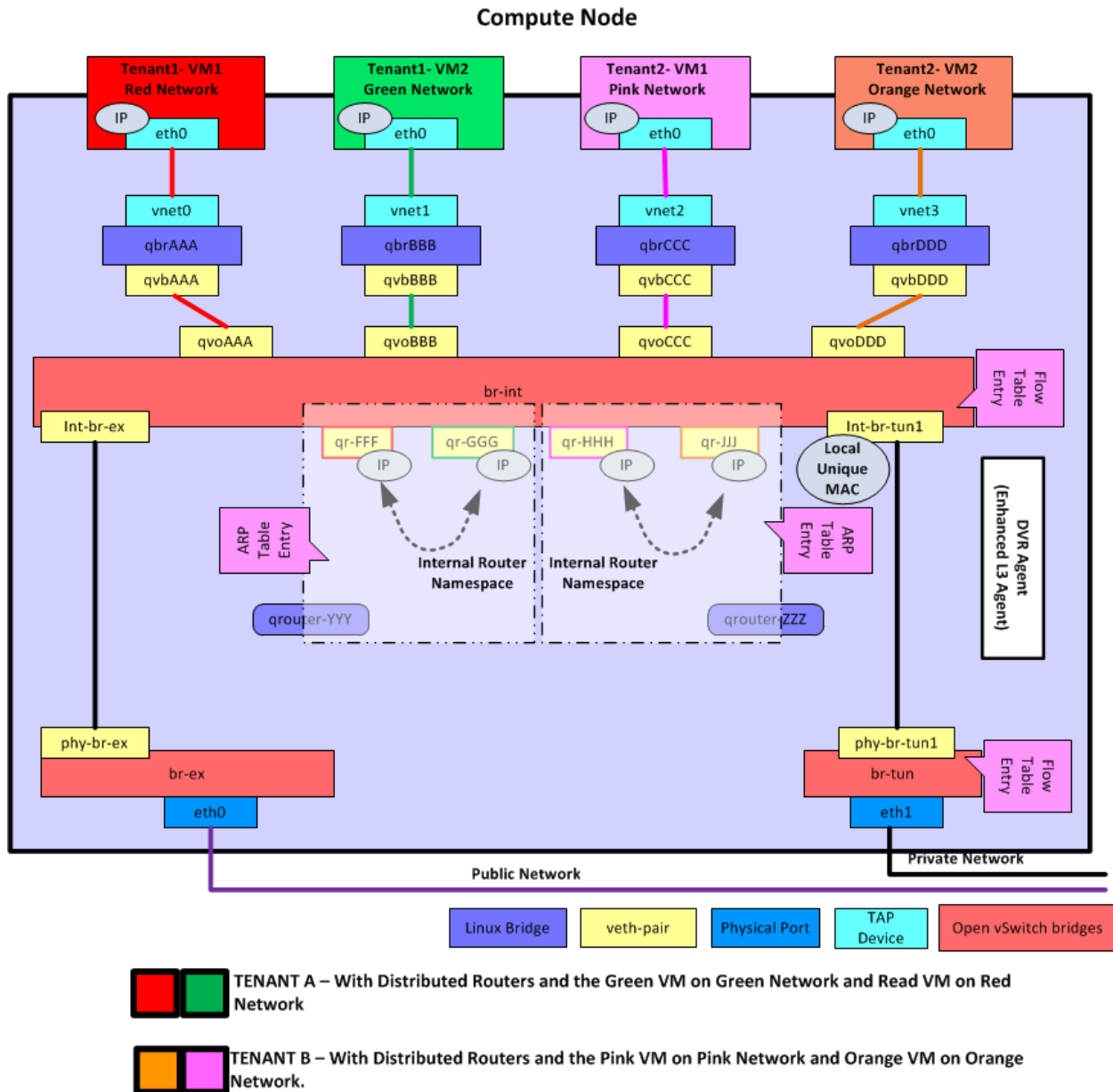
FIP - Floating IP

SNAT - Source NAT

Detailed Architecture Overview:



IR (Internal Router) Namespace within a compute Node for two Tenants.

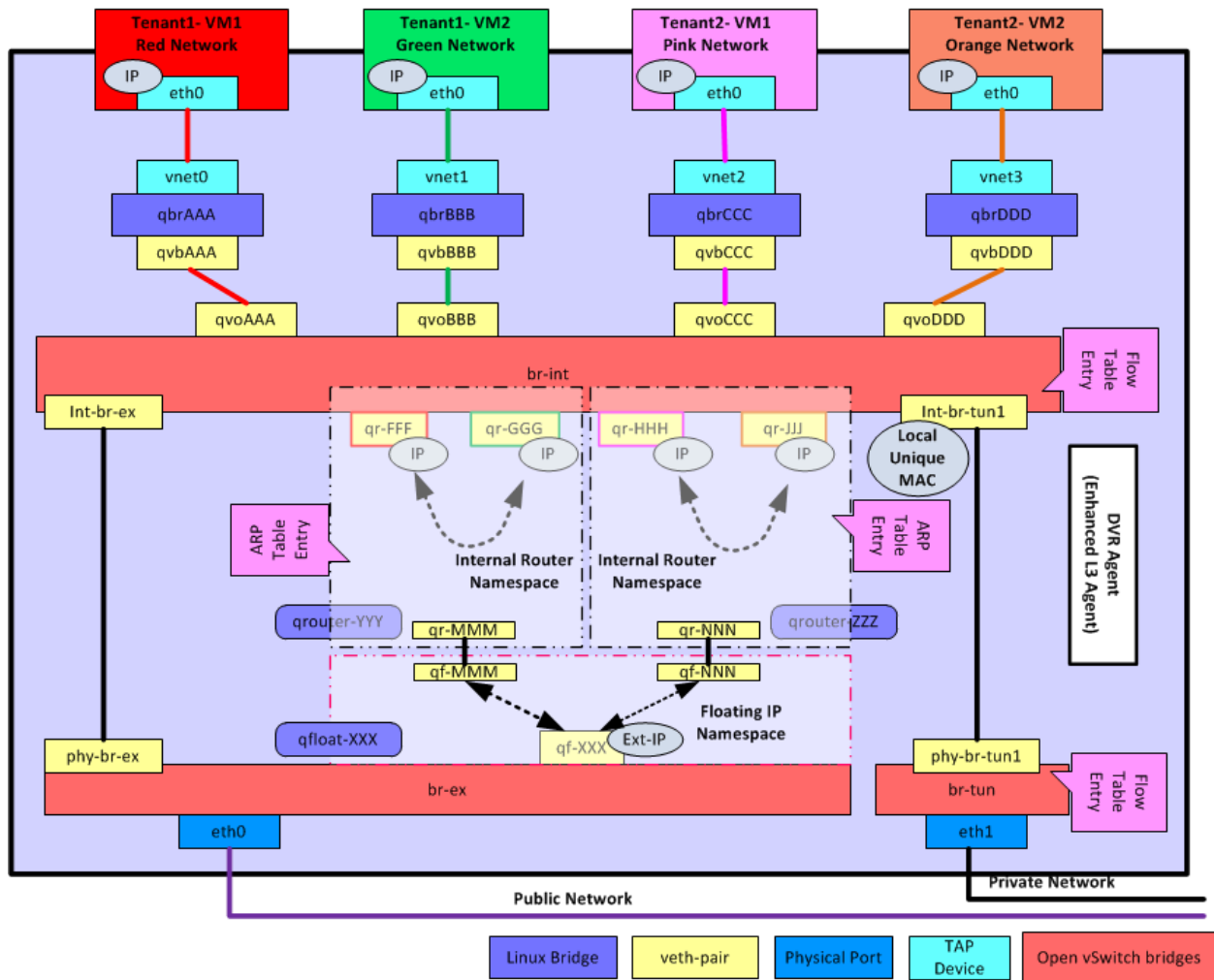


The picture above shows how the Internal Routers (IR) are created in the Compute Node. Each Router will run on its own namespace. The DVR Agent or the Enhanced L3 Agent is responsible for create/delete/update the routers in the router namespace and also to populate the ARP entry for all clients in the network owned by the router.

NOTE: Even though there is a connection between the int-br-ex and phy-br-ex in the picture the actual traffic will flow through the namespace and not through this link.

Namespace split between the IR and the Floating IP in the Compute Node.

Compute Node



- TENANT A – With Distributed Routers and the Green VM on Green Network and Red VM on Red Network
- TENANT B – With Distributed Routers and the Pink VM on Pink Network and Orange VM on Orange Network.

Z

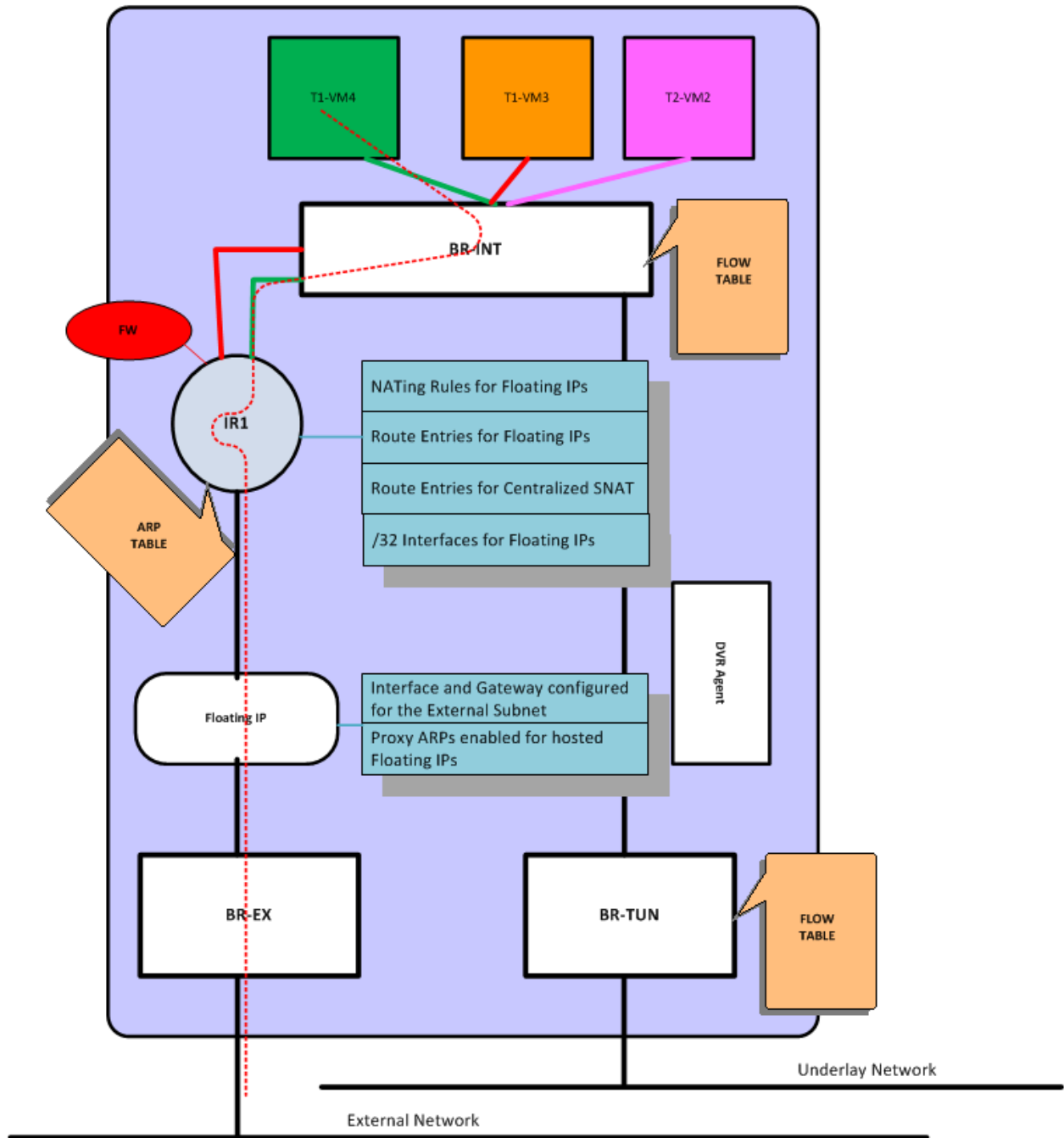
The picture above shows how the Floating IPs are associated to the VMs in the Compute Node. Each IR Router will run on its own namespace and also the Floating IP will run on its own namespace. The DVR Agent or the Enhanced L3 Agent is responsible for create/delete/update the routers in the router namespace and also to populate the ARP entry for all clients in the network owned by the router. Also the same DVR Agent will be responsible for creating the Floating IP Namespace and forwarding the traffic.

NOTE: Even though there is a connection between the int-br-ex and phy-br-ex in the picture the

actual traffic will flow through the namespace and not through this link.

The picture below shows at a high level what need to be done for the Floating IP to work right in the Compute Node.

Floating IP Traffic Flow and Programming



Plugin/Extension:

1. Split the External Gateway from the Router Object

Table Definition

1.1.1.1.1 Router Table Extension

Attribute	Type	Required	CRUD	Default Value	Validation Constraint
Distributed	bool	No	CRUD	False	True/False

Table 1: Router Table Extension

Agent:

The current proposal is we would be using the L3 Agent in different modes based on the config file option. Either it can be a legacy L3 Agent or Enhanced L3 Agent running in DVR mode.

Single enhanced l3-agent/scheduler. All new functionality/roles would be enabled from l3_agent.ini flags. The new flags would be sent to the plugin via configuration fields in the agent status dict. If none of the flags are set, the l3-agent would act like a legacy agent. It should be possible to have a legacy centralized router and dvr-IR on the same node being managed by a single agent process.

- o Advantages

- Same deployment – no new processes/agents to HA

- Maximize code re-use – existing code can be leveraged/enhanced for new functionality

- Scheduler can identify agents functionality from the configuration fields

- Less impact to/from “agent consolidation” movement in OpenStack

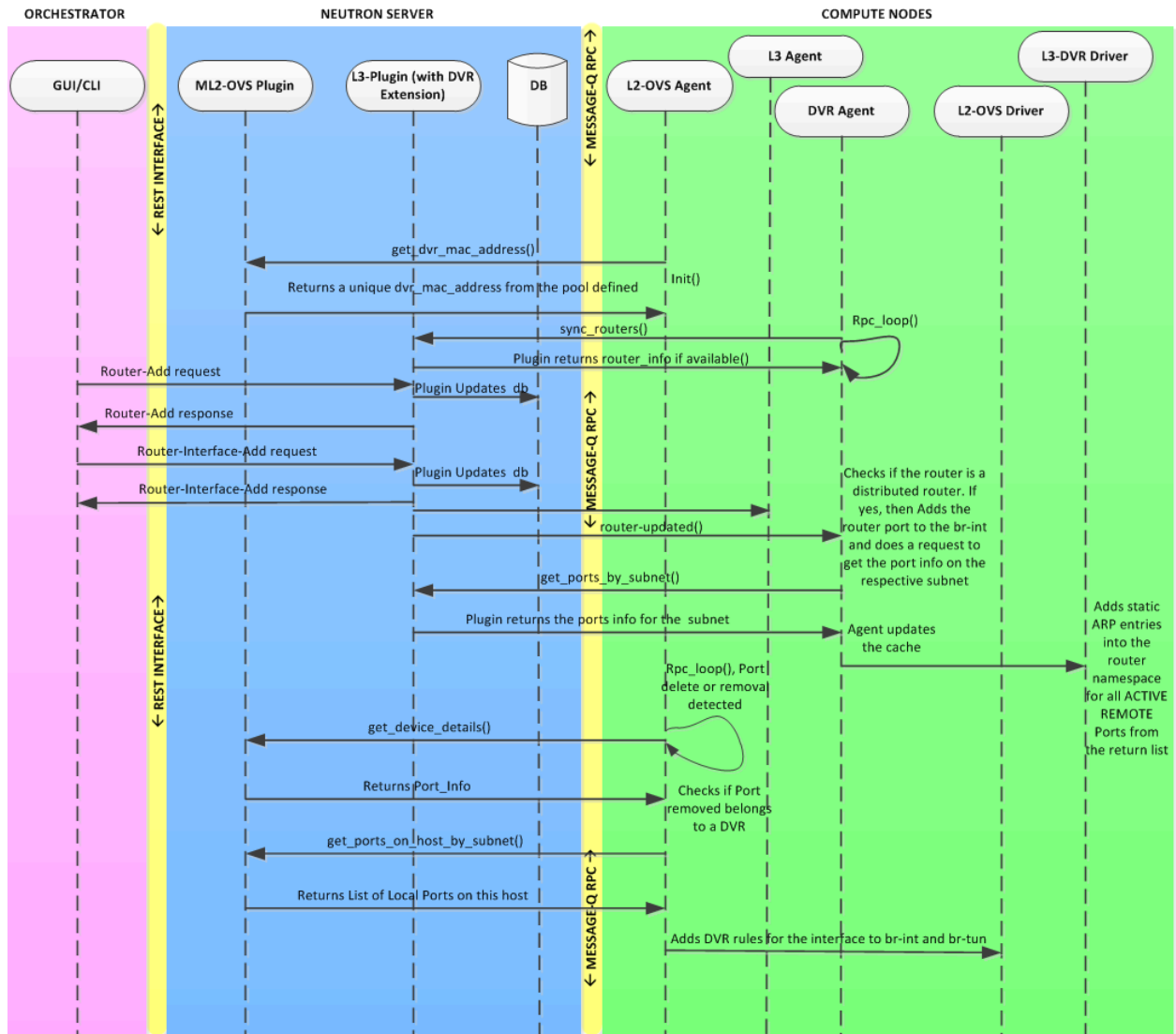
L2 Changes

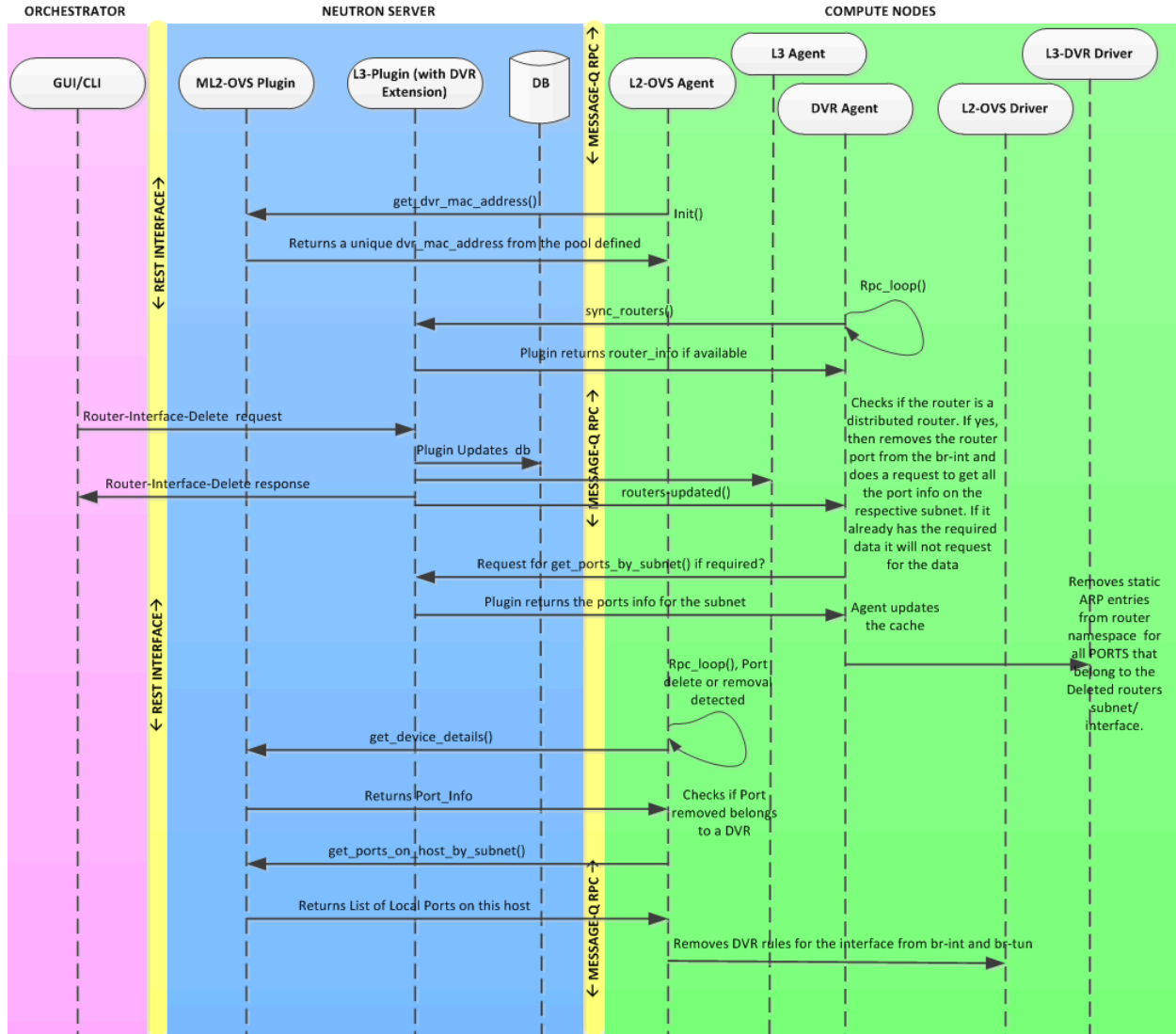
Even though we are working on the Distributed L3 Agent which is an L3 Service there are some dependencies that are related to the ML2 Plugin. So there is some work in the ML2 Plugin and Agent that would provide the L2 Level Flow configuration for the Distributed Virtual Router.

DVR Agent - ARP Proxy

The DVR Agent will be responsible for populating the ARP Table in the router namespace. So no ARP packets will be forwarded from the Router.

ML2/L3 Plugin communication with the Agents (Flow diagram).





API/CLI:

We are proposing a set of API's (ADMIN ONLY) that would help admins to do the actual migration from the legacy model to the new distributed model. These API's cannot be used by the Tenant.

router-create **Create a router for a given tenant.**

Admin can only set this attribute. The tenant's need not be aware about this attribute in the router table. So it is not visible to the tenant.

Request

```
POST /v2.0/routers
Accept: application/json
```

```
{
"router":
{
"name": "another_router",
"admin_state_up": true,
"distributed": true
}
}
```

Response

```
{
"router":
{
"status": "ACTIVE",
"external_gateway_info": null,
"name": "another_router",
"admin_state_up": true,
"distributed": true,
"tenant_id": "6b96ff0cb17a4b859e1e575d221683d3",
"id": "8604a0de-7f6b-409a-a47c-a1cc7bc77b2e"
}
}
```

router-list-on-l3-agent **List the routers on a L3 agent.**

No Change

router-port-list **List ports that belong to a given tenant, with specified router.**

No Change

router-show **Show information of a given router.**

Request

```
GET /v2.0/routers/a9254bdb-2613-4a13-ac4c-adc581fba50d
Accept: application/json
```

Response

```
{
"routers":
[
{
"status": "ACTIVE",
"external_gateway_info":
{
"network_id": ""
},
"name": "router1",
```

```
"admin_state_up": true,
"distributed": true,
"tenant_id": "33a40233088643acb66ff6eb0e6ea679",
"id": "a9254bdb-2613-4a13-ac4c-adc581fba50d"

}]
}
```

router-update **Update router's information.**

Request

```
PUT /v2.0/routers
Accept: application/json/8604a0de-7f6b-409a-a47c-a1cc7bc77b2e
{
  "router":
  {
    "name": "new_router",

    "external_gateway_info":
    {
      "network_id": ""
    }
  }
}
```

Response

```
{
  "router":
  {
    "status": "ACTIVE",
    "external_gateway_info":
    {
      "network_id": ""
    },
    "name": "new_router",
    "admin_state_up": true,
    "distributed": true,
    "tenant_id": "6b96ff0cb17a4b859e1e575d221683d3",
    "id": "8604a0de-7f6b-409a-a47c-a1cc7bc77b2e"
  }
}
```

Additional API

There will be API to show the “SNAT agent association” and to associate a SNAT with a different

agent. This would be useful for Admins to move around the SNAT or provide a different agent affinity.

<code>l3-agent-list-hosting-snat</code>	List L3 agents hosting snat.
<code>l3-agent-snat-add</code>	Add a snat to a L3 agent.
<code>l3-agent-snat-remove</code>	Remove a snat from a L3 agent.

CLI Commands

Router Create:

NOTE: Distributed Flag can be also set using the config file. The setting in the API or CLI will override the config file setting.

For distributed router. (ADMIN Only command)

```
neutron router-create router1 --distributed=true
```

For Centralized router. (ADMIN Only command)

```
neutron router-create router1 --distributed=False
```

```
neutron router-create router1 ( In this case, if the API does not have the distributed flag, then the setting in the config file will be used).
```

Admin only CLI commands for

```
l3-agent-list-hosting-snat List L3 agents hosting a snat service.
```

This command will list the agent with the router-id and SNAT IP.

```
l3-agent-snat-add Add a snat association to an L3 agent.
```

This command will allow an admin to associate a SNAT service to a router. This command will take the router ID as an argument.

```
l3-agent-snat-remove Remove snat association from an L3 agent.
```

This command will allow an admin to remove or disassociate a SNAT service from the agent.

Links:

https://docs.google.com/drawings/d/1GGwbLa72n8c2T3SBApKK7uJ6WLTSRa7erTI_3QNj5Bg/edit

<https://blueprints.launchpad.net/neutron/+spec/l3-agent-consolidation>