



redhat



eNovance  
from Red Hat®

# High Availability in Neutron

Getting the L3 Agent Right

**Sylvain Afchain**  
Software Engineer

**Assaf Muller**  
Software Engineer

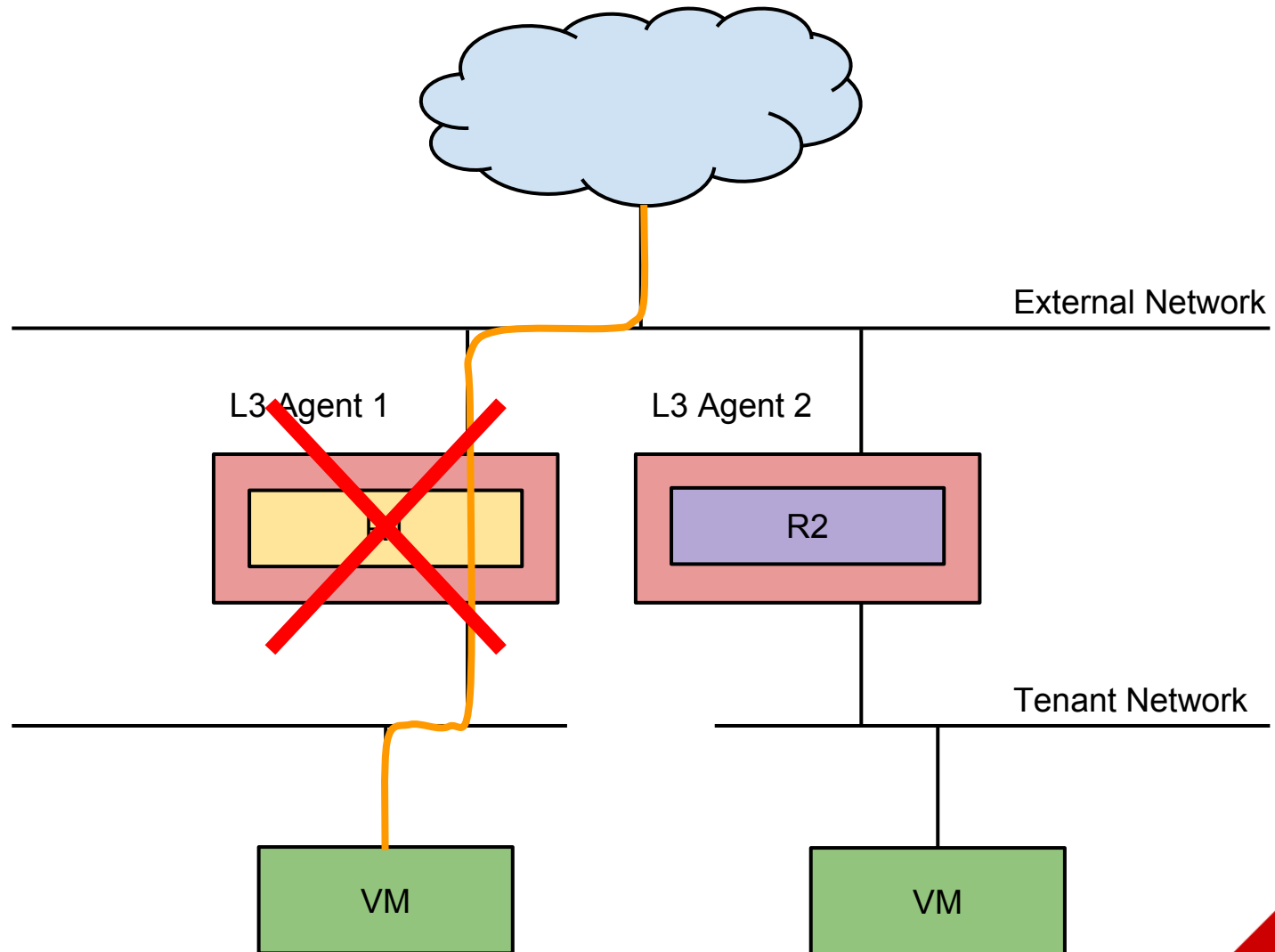
November 2014



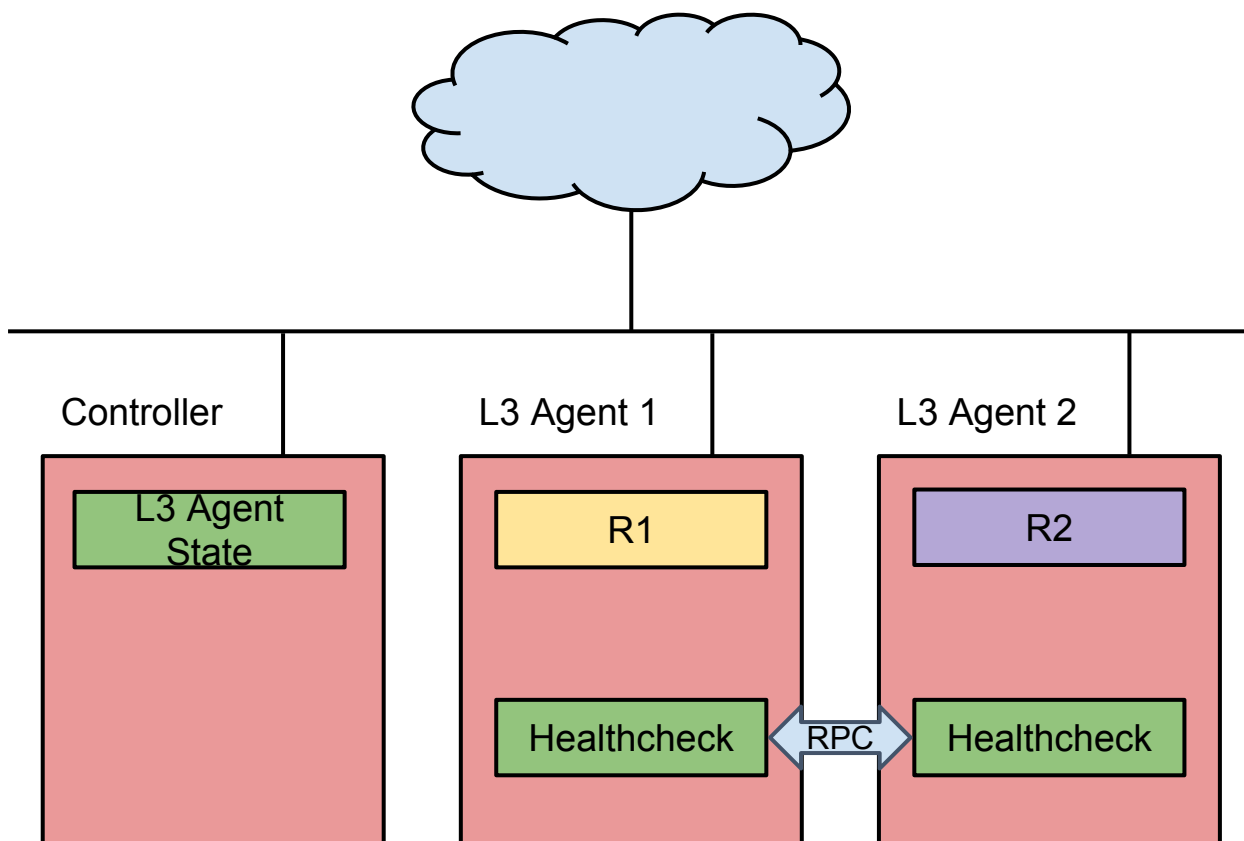
# SITUATION IN ICEHOUSE



# Single Point of Failure

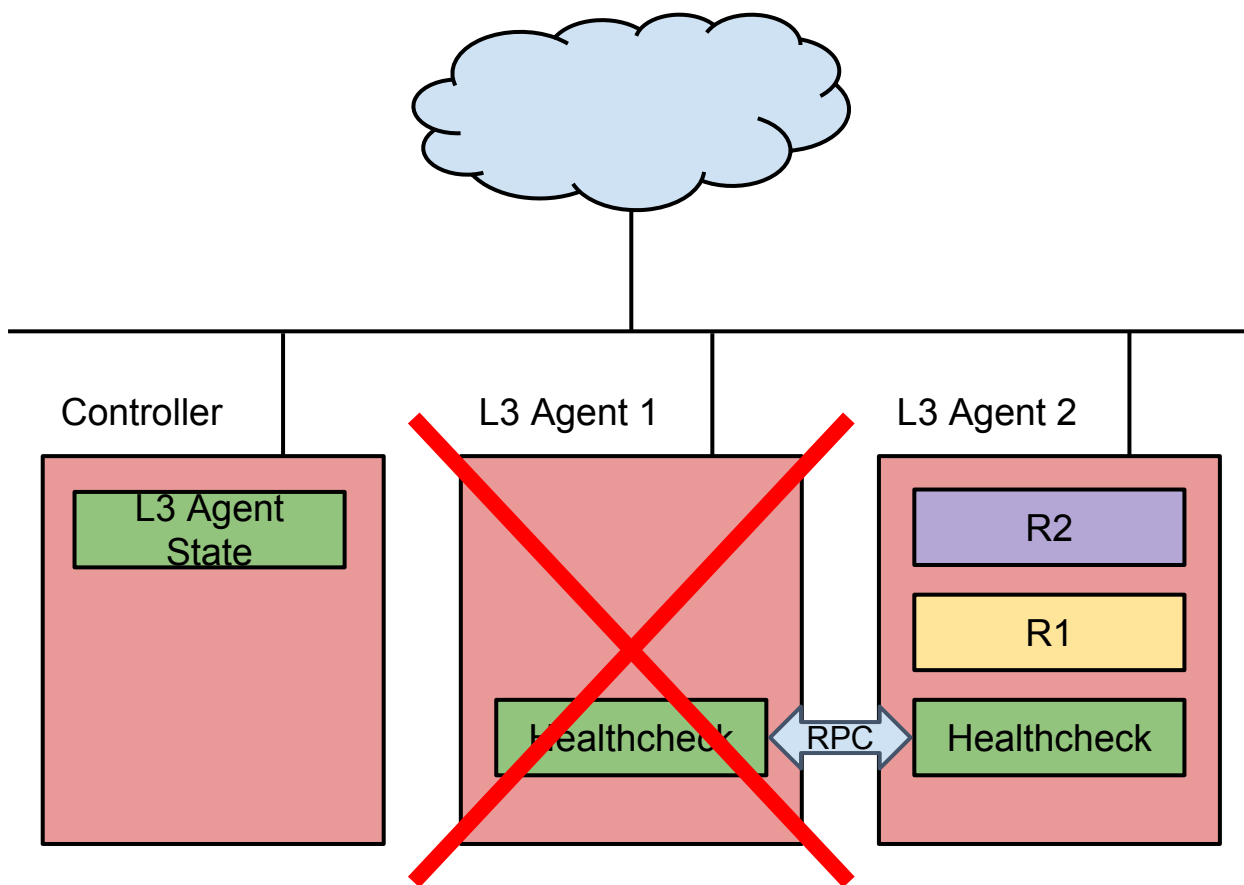


# L3 Agent Healthcheck



- Developed by eNovance
- Open-Source
- Works and tested on Grizzly, Havana and Icehouse

# L3 Agent Healthcheck



- Developed by eNovance
- Open-Source
- Works and tested on Grizzly, Havana and Icehouse



# L3 Agent Healthcheck

## Pros

- Does not affect deployment
- Remove a node if isolated
- Distributed service
- Works since Grizzly
- Lightweight

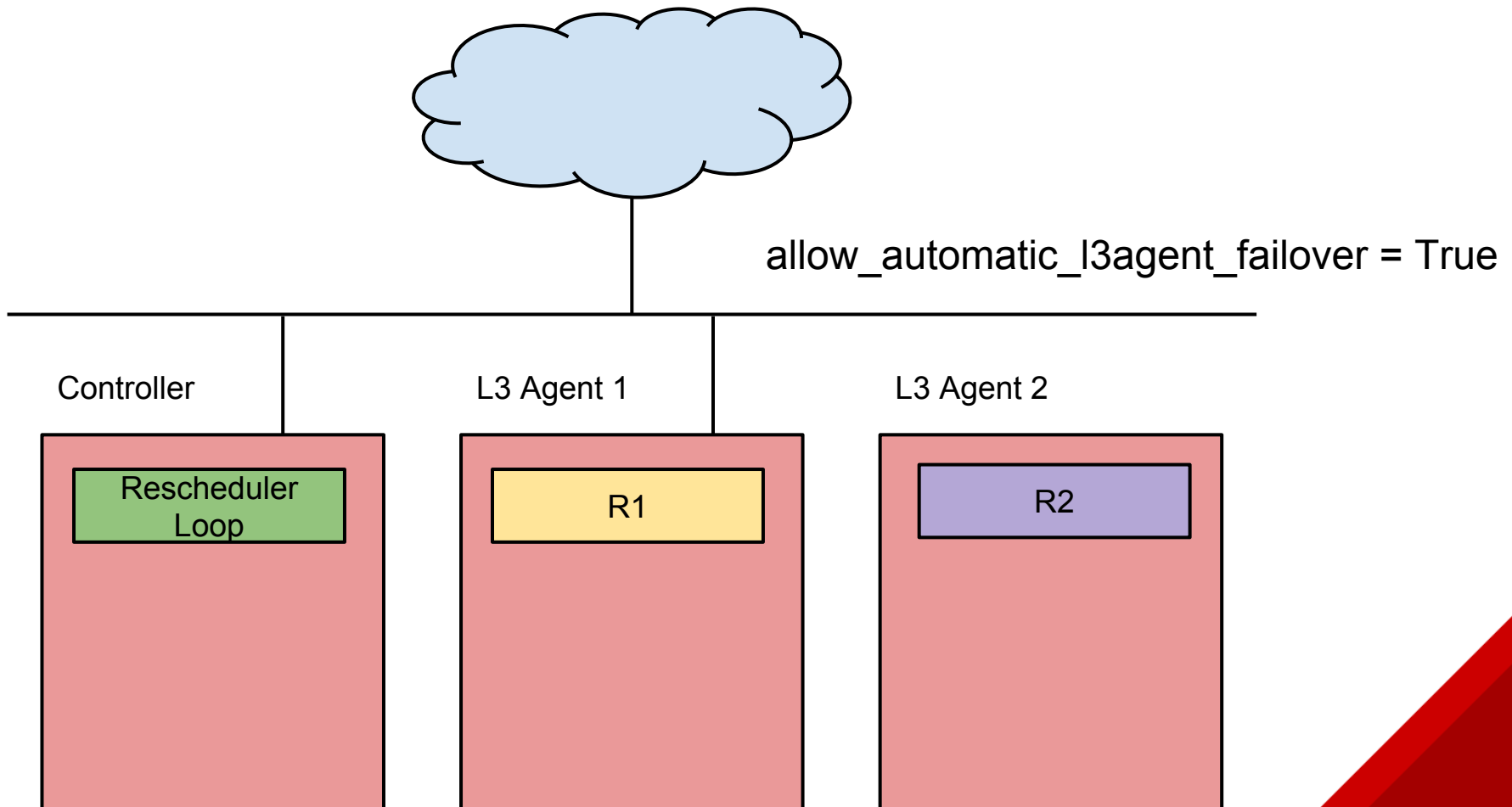
## Cons

- Still no full HA
- Not stateful
- Long downtime
- Out of tree
- Still not the right way

# SITUATION IN JUNO

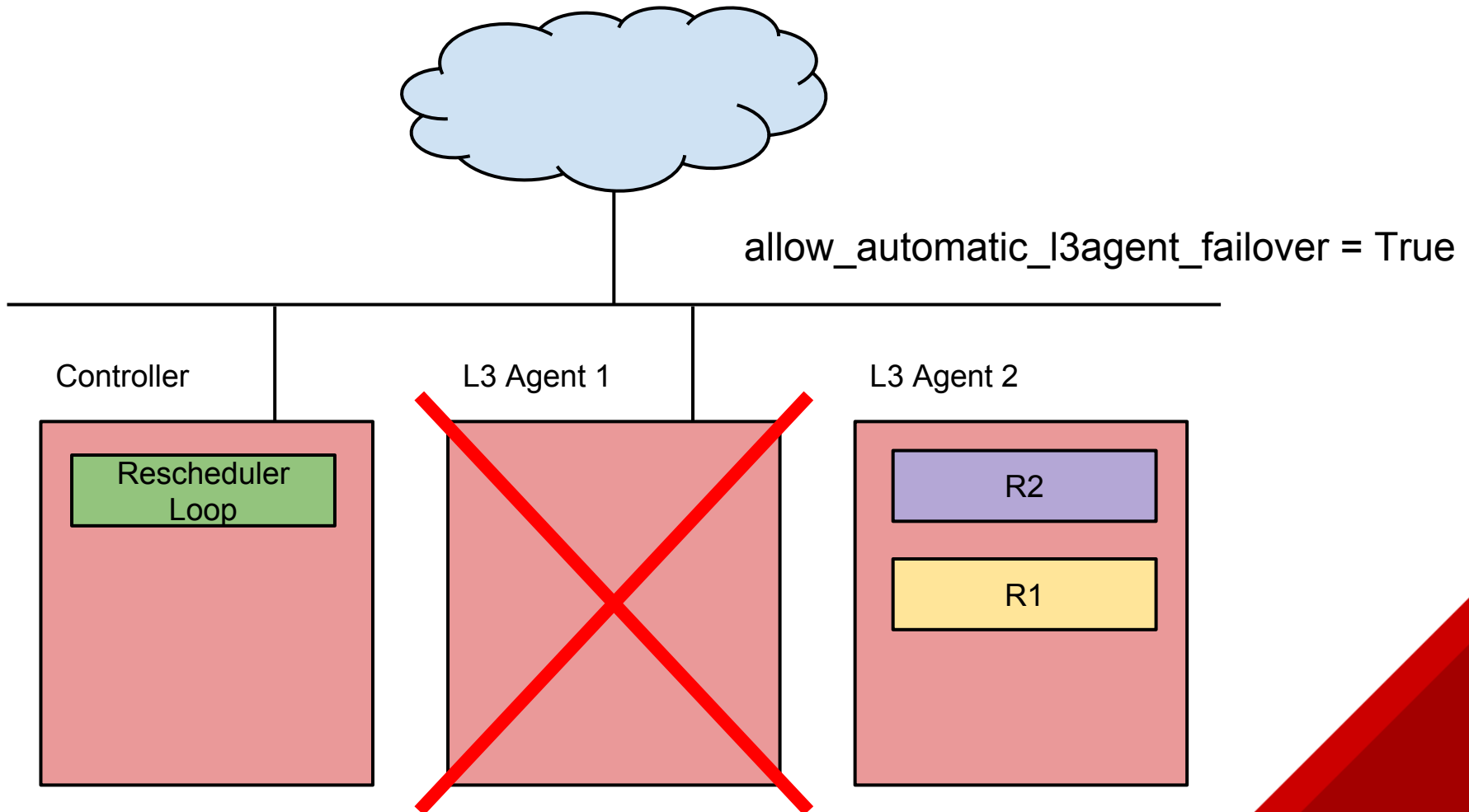


# Controller Rescheduling



Authored by Kevin Benton

# Controller Rescheduling

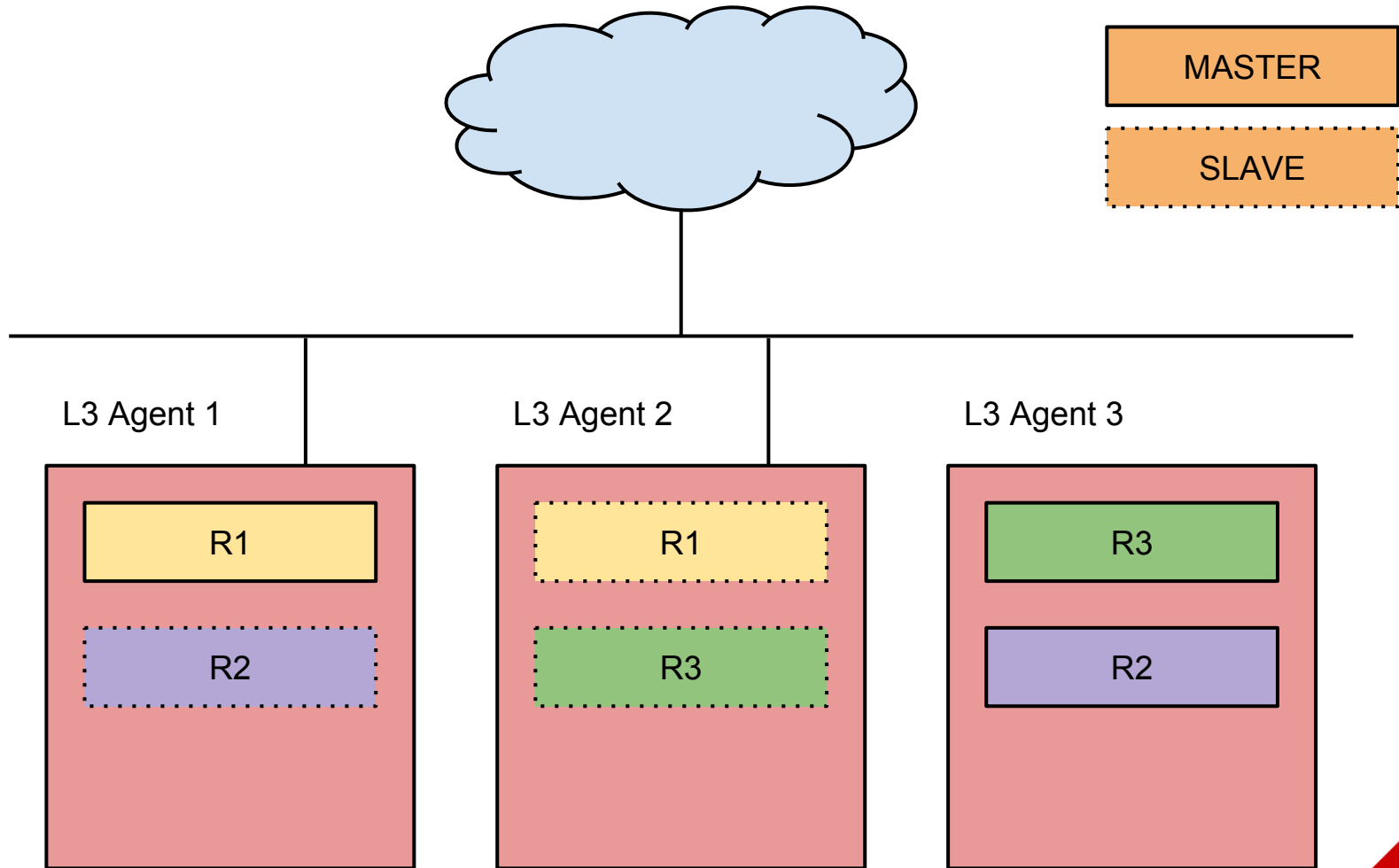


# Another Approach: Keepalived

## Features

- Configuration determines default, admin can overrule
- Works within tenant networks
- Plays nice with: FWaaS, VPNaaS, LBaaS
- Failover independent from RPC layer
- Routers in active / passive
- Floating IP in active / passive
- All L3 agents are active
- Uses keepalived (VRRP)
- Rough failover time:  
 $c + 2/29 * n$

# VRRP: Pre-emptive Router Scheduling



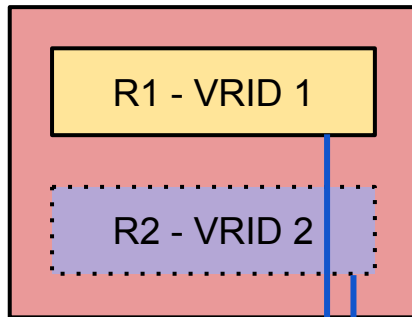


# VRRP: HA Networks & VRID

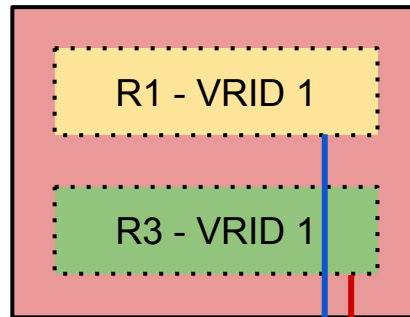
R1, R2 - Tenant A

R3 - Tenant B

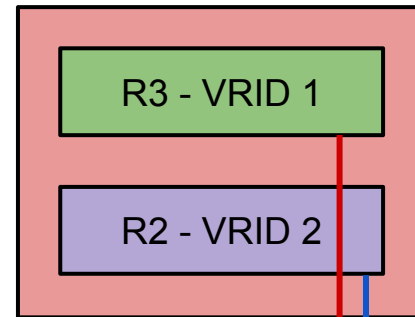
L3 Agent 1



L3 Agent 2



L3 Agent 3



# VRRP: Implementation

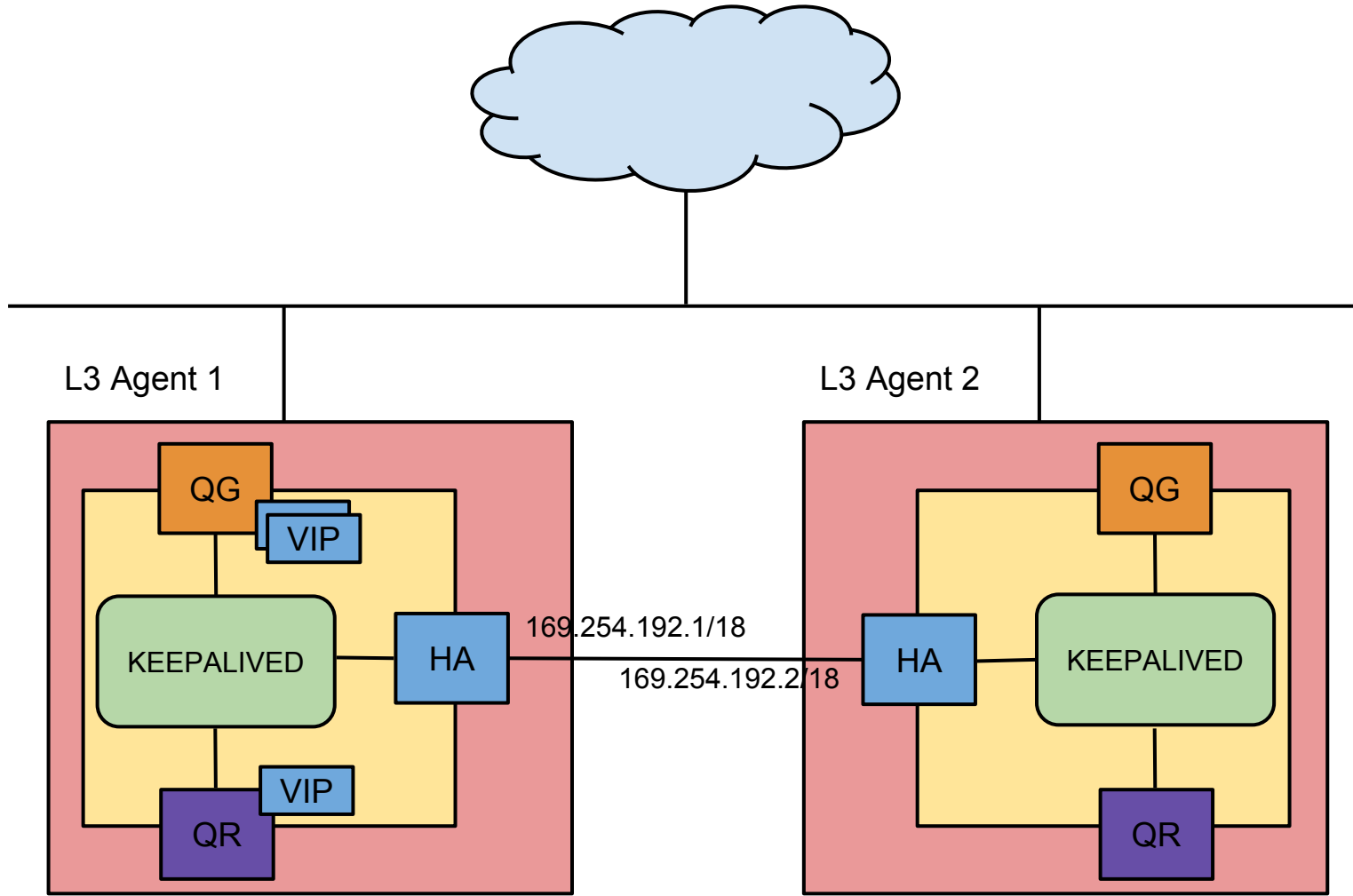
## Neutron Server

- Per tenant network to accommodate VRRP traffic
- New virtual router ID attribute uniquely identifies router clusters

## L3 Agent

- New keepalived manager
- IPs as VIPs, only present on the master instance

# VRRP: Implementation



# VRRP: CLI/API

## neutron net-list

As an admin :

id	name	subnets
27b2e2b7-22c0-4f74-	private	f472d8c8-1dd0-4272- 10.0.0.0/24
be1a07de-9d7b-4823-	public	7f3d69a6-bd50-4cd5- 172.24.4.0/24
62917a72-0576-422e-	HA network tenant	12f466f4-6c51-4726- 169.254.192.0/18

As a user :

id	name	subnets
27b2e2b7-22c0-4f74-	private	f472d8c8-1dd0-4272- 10.0.0.0/24
be1a07de-9d7b-4823-	public	7f3d69a6-bd50-4cd5- 172.24.4.0/24



# VRRP: CLI/API

## neutron port-list

As an admin :

```
+-----+-----+-----+
| id          | name    | fixed_ips                                     |
+-----+-----+-----+
| 3e7268c6-   | HA port | {"subnet_id": "12f466f4-", "ip_address": "169.254.192.2"} |
| d87a6c9c-   | HA port | {"subnet_id": "12f466f4-", "ip_address": "169.254.192.1"} |
| 5105bd78-   |         | {"subnet_id": "f472d8c8-", "ip_address": "10.0.0.1"}      |
+-----+-----+-----+
```

As a user :

```
+-----+-----+-----+
| id          | name    | fixed_ips                                     | +-----+
+-----+-----+-----+
| 5105bd78-   |         | {"subnet_id": "f472d8c8-", "ip_address": "10.0.0.1"}      |
+-----+-----+-----+
```

# VRRP: CLI/API

Where are my HA router instances?

```
neutron l3-agent-list-hosting-router router1
```

```
+-----+-----+-----+-----+
| id          | host    | admin_state_up | alive |
+-----+-----+-----+-----+
| 0dad7203-cba8-4b79-bec3-16ddf55d6a5a | ops-1   | True           | :-)  |
| 7d7afb99-b522-442a-8acd-f1548a1dea19 | ops-2   | True           | :-)  |
+-----+-----+-----+-----+
```



# VRRP : Future Work

## Improvements

- Administration:
  - Where is the master?  
Can I move it?
  - Log state transitions
- Conntrackd for stateful SNAT traffic
- Migrate legacy routers to HA
- In Juno, a router may be distributed or HA, but not both



# VRRP: Limitations

## Limitations

- 255 virtual routers per HA network thus per tenant
- Can be removed by allowing more than one HA network per tenant

## Improvements

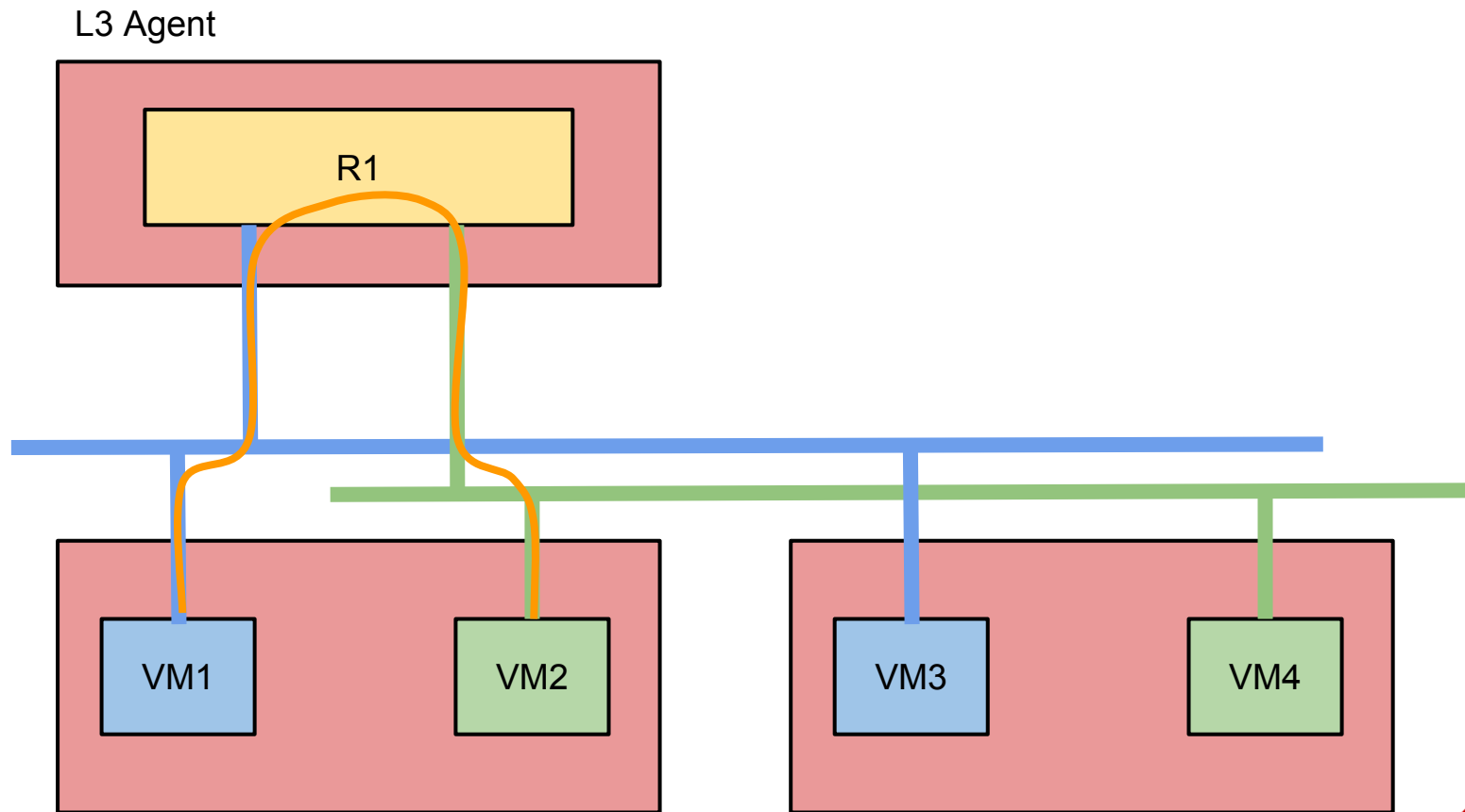
- East-west traffic could be improved, no need to go through a L3 node
- North-south traffic could be improved as well, especially for floating IPs

# Another Approach: DVR

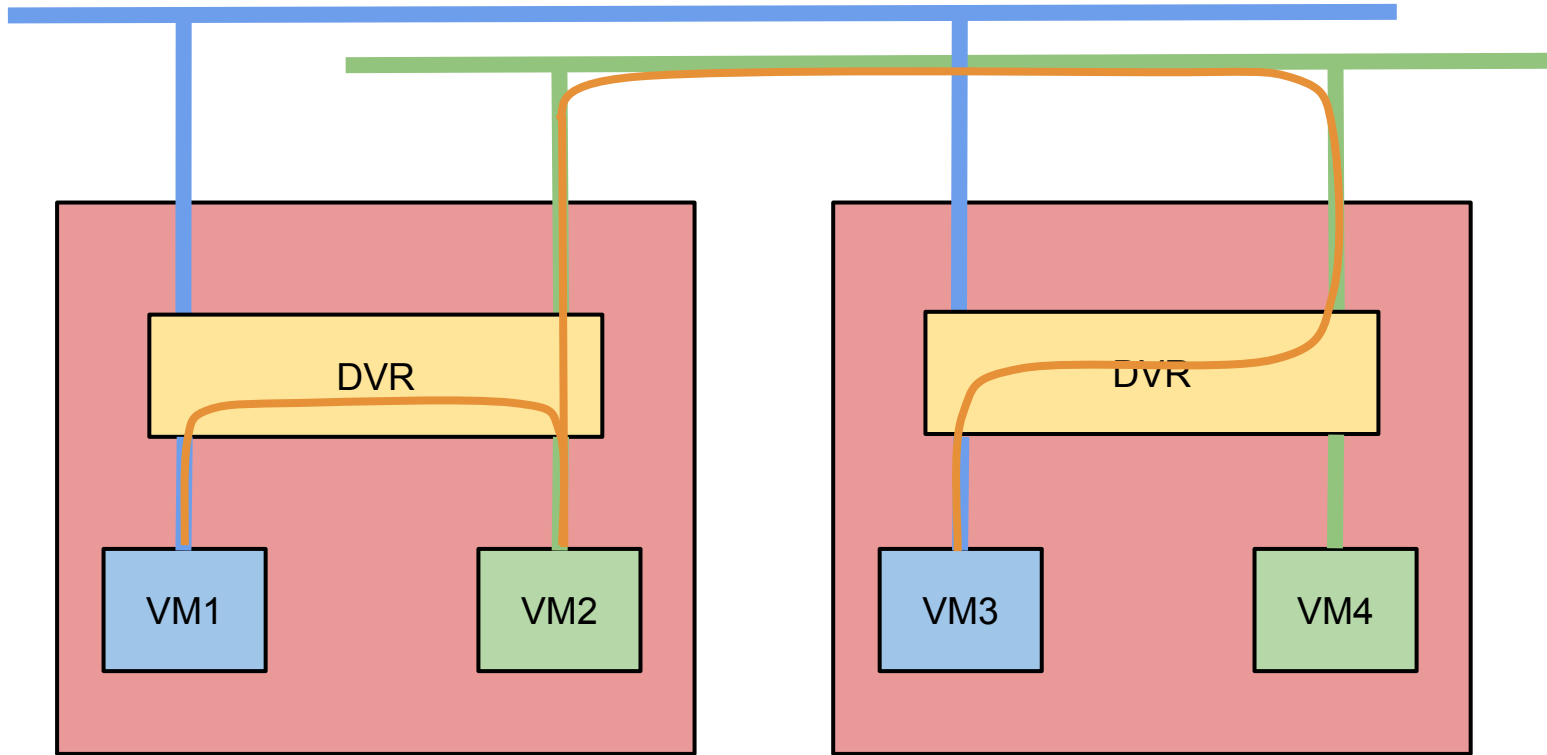
## Features

- Distributes virtual routers on all compute nodes
- No more east-west traffic through L3 node
- No single point of failure
- Floating IPs hosted by the compute node hosting the VM
- Some service could be distributed as well, like FWaaS
- SNAT support through a L3 node

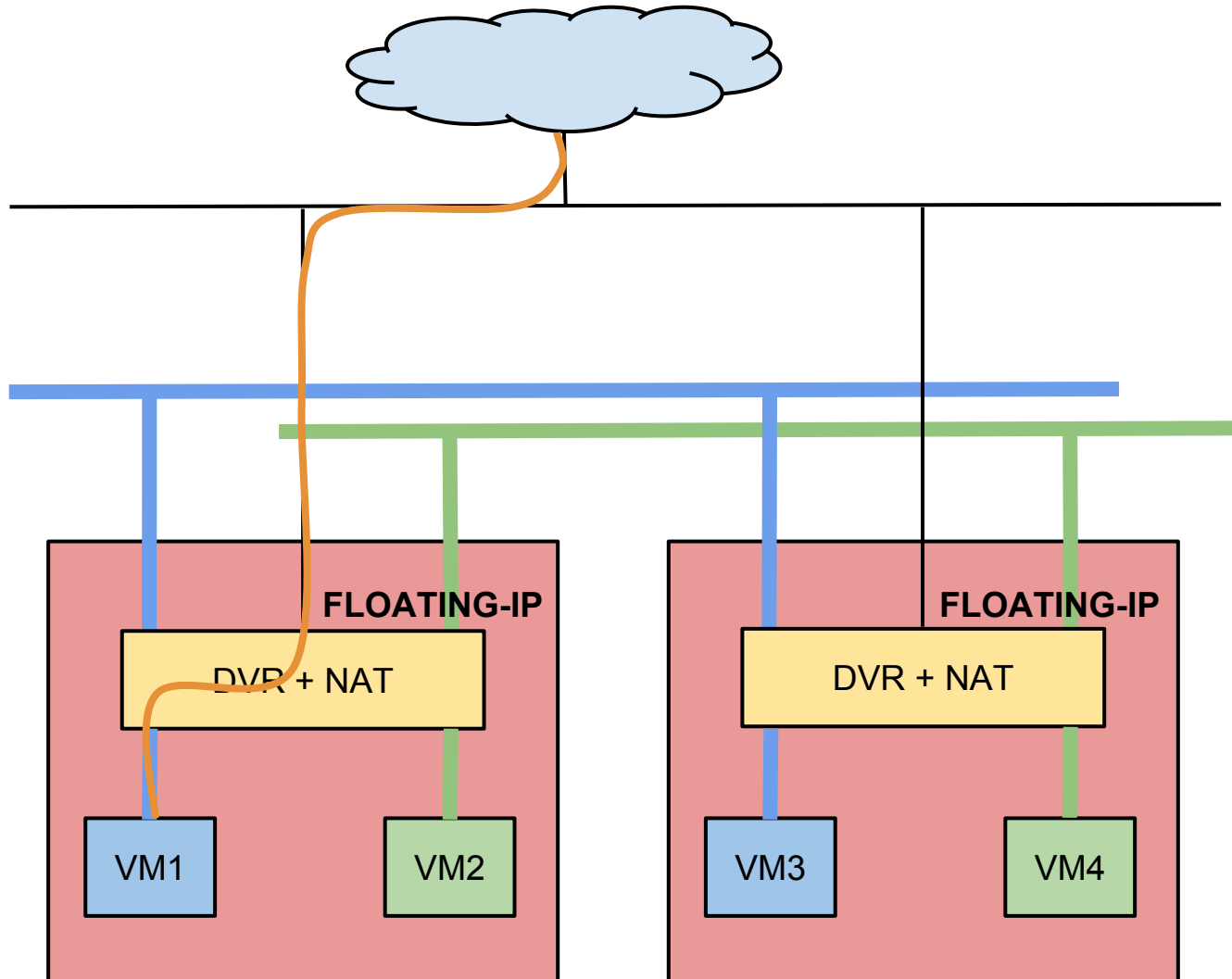
# Without DVR



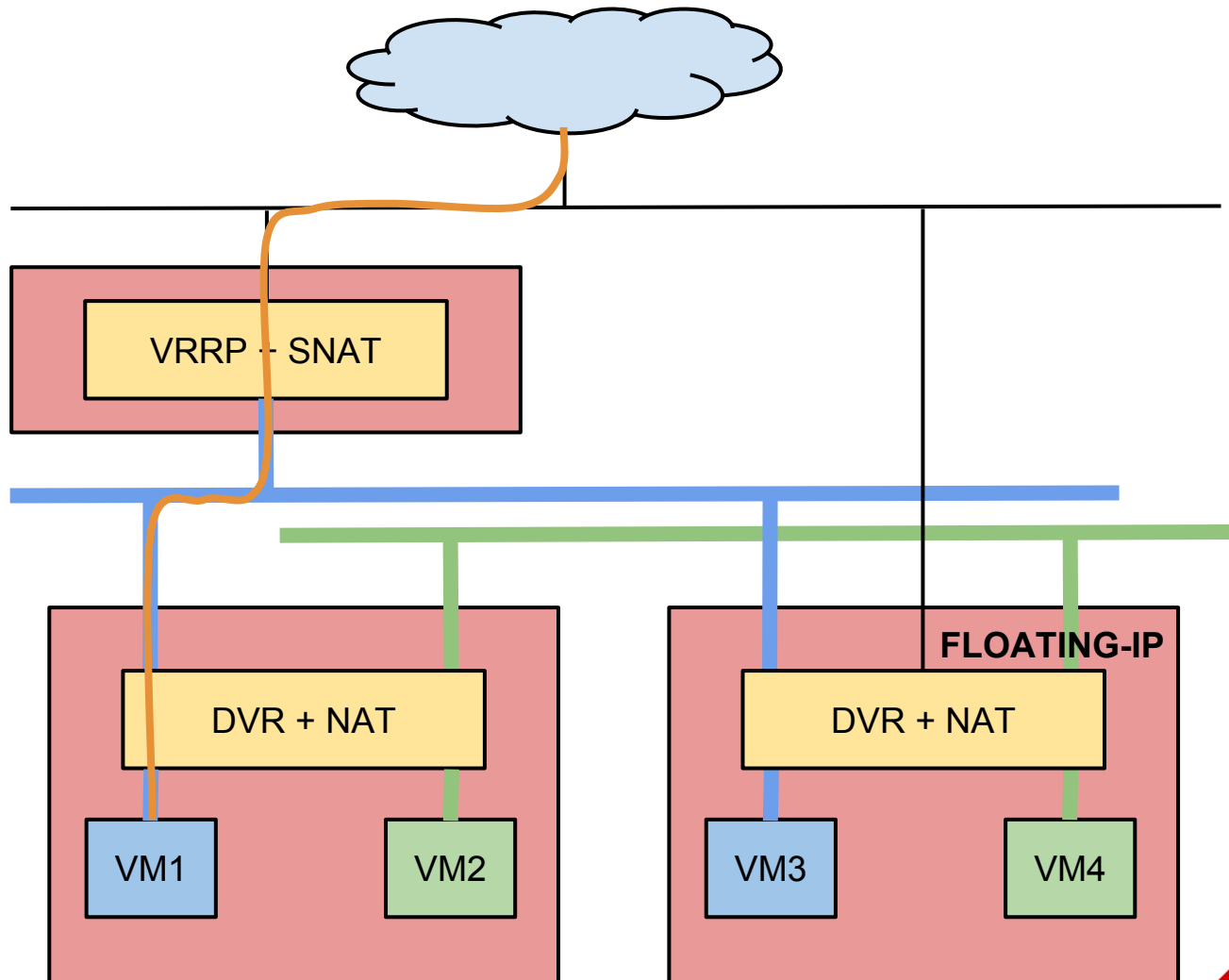
# East-West with DVR



# Floating IPs with DVR



# SNAT with DVR



# Summary

	Healthcheck	Rescheduling	L3 HA	DVR
Advantages	Mature	In-tree, simple	Failover is quick and indifferent to management plane	Removes bottleneck from network node
Release	Grizzly	Juno	Juno	Juno
Segmentation Technology	All	All	All	Tunneling + L2pop
Topology	Extra agent to install on network nodes	Enable configuration option	Enable configuration option	Compute nodes connected to external network(s)



**redhat.**



**enovance**  
from Red Hat®

**Assaf Muller**

Software Engineer

[amuller@redhat.com](mailto:amuller@redhat.com)

amuller on Freenode

[assafmuller.com](http://assafmuller.com)

**Sylvain Afchain**

Software Engineer

[sylvain.afchain@enovance.com](mailto:sylvain.afchain@enovance.com)

safchain on Freenode

[Enovance Tech Notes](#)