



Trellis Tutorial

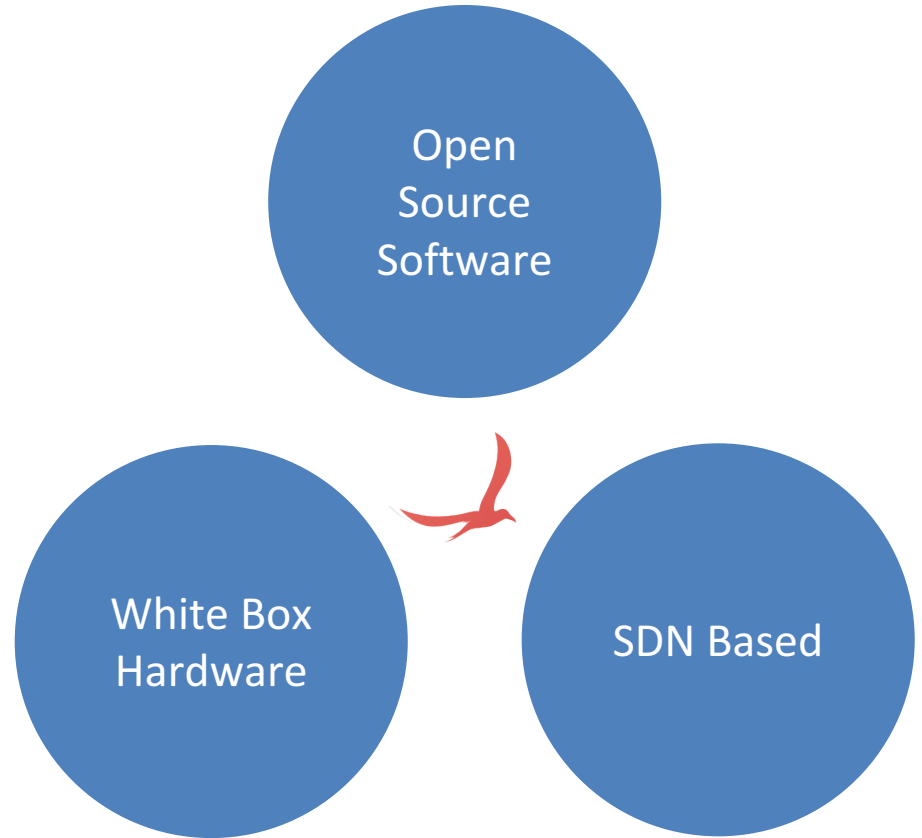
An Open-Source, White-Box, SDN Based Leaf-Spine Network Fabric

Charles Chan

Open Networking Foundation

Trellis

Multi-purpose leaf-spine
fabric designed for NFV



Trellis Features

Bridging

Routing

VLAN

L3 DHCP
Relay

vRouter

IPv6

Dual
Homing

and more...

Prerequisite

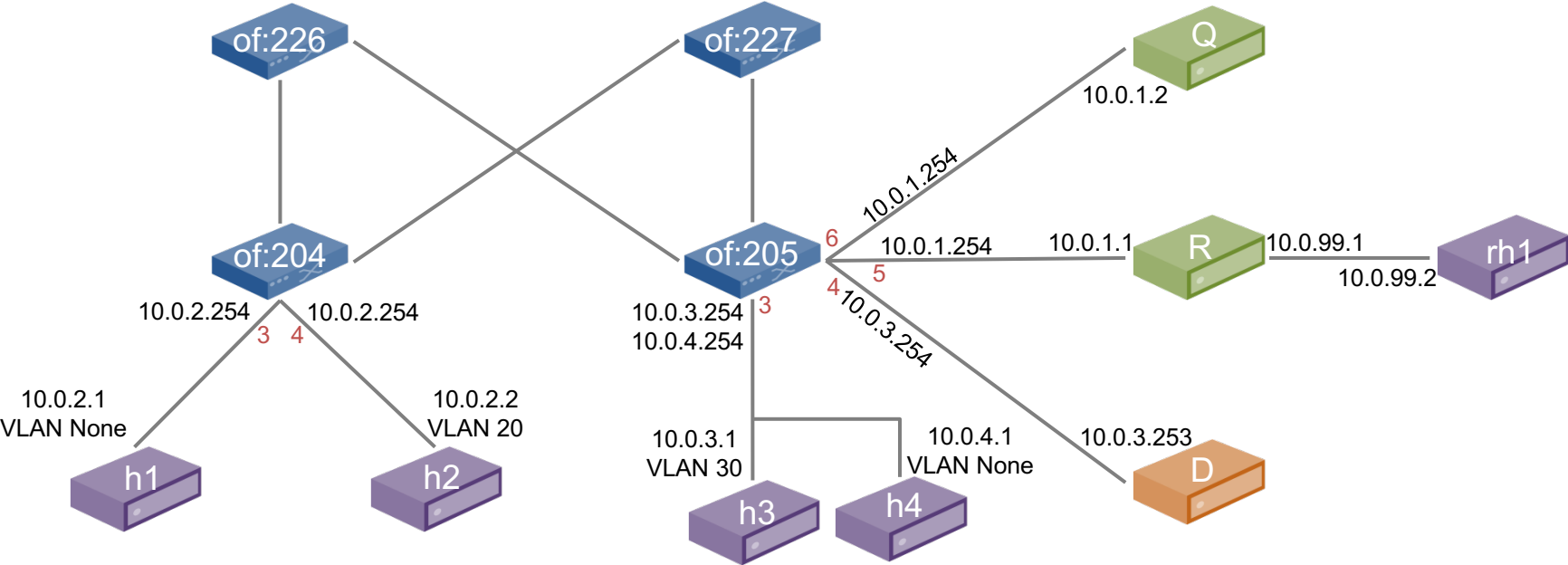
ONOS

Mininet

DHCP
Server

Quagga
&
BGP

Demo Topology



Q: Quagga, R: Upstream Router, D: DHCP Server



Device

```
{
  "devices" : {
    "of:00000000000000204" : {
      "segmentrouting" : {
        "name" : "s204",
        "ipv4NodeSid" : 204,
        "ipv4Loopback" : "192.168.0.204",
        "routerMac" : "00:00:00:00:02:04",
        "isEdgeRouter" : true,
        "adjacencySids" : []
      },
      "basic" : {
        "name": "s204",
        "driver" : "ofdpa-ovs"
      }
    }
  }
}
```

Data Path ID (DPID) of the switch

Device

```
{
  "devices" : {
    "of:00000000000000204" : {
      "segmentrouting" : {
        "name" : "s204",
        "ipv4NodeSid" : 204,
        "ipv4Loopback" : "192.168.0.204",
        "routerMac" : "00:00:00:00:02:04",
        "isEdgeRouter" : true,
        "adjacencySids" : []
      },
      "basic" : {
        "name": "s204",
        "driver" : "ofdpa-ovs"
      }
    }
  }
}
```

Unused. This name doesn't really matter.
Will be deprecated.

Device

```
{
  "devices" : {
    "of:00000000000000204" : {
      "segmentrouting" : {
        "name" : "s204",
        "ipv4NodeSid" : 204,
        "ipv4Loopback" : "192.168.0.204",
        "routerMac" : "00:00:00:00:02:04",
        "isEdgeRouter" : true,
        "adjacencySids" : []
      },
      "basic" : {
        "name": "s204",
        "driver" : "ofdpa-ovs"
      }
    }
  }
}
```

Segment ID.

Can be an arbitrary value but need to be unique in the system.

Also used as the MPLS label when doing segment routing.

Do not use reserved MPLS labels (i.e. ≤ 16).

Device

```
{
  "devices" : {
    "of:00000000000000204" : {
      "segmentrouting" : {
        "name" : "s204",
        "ipv4NodeSid" : 204,
        "ipv4Loopback" : "192.168.0.204",
        "routerMac" : "00:00:00:00:02:04",
        "isEdgeRouter" : true,
        "adjacencySids" : []
      },
      "basic" : {
        "name": "s204",
        "driver" : "ofdpa-ovs"
      }
    }
  }
}
```

Loopback address of the switch.

Can be an arbitrary value (e.g. the management IP) but need to be unique in the system.

Device

```
{
  "devices" : {
    "of:00000000000000204" : {
      "segmentrouting" : {
        "name" : "s204",
        "ipv4NodeSid" : 204,
        "ipv4Loopback" : "192.168.0.204",
        "routerMac" : "00:00:00:00:02:04",
        "isEdgeRouter" : true,
        "adjacencySids" : []
      },
      "basic" : {
        "name": "s204",
        "driver" : "ofdpa-ovs"
      }
    }
  }
}
```

Use for ARP reply when host ARPs its gateway.
Can be an arbitrary value (e.g. the management MAC) but need to be unique in the system.

Device

```
{
  "devices" : {
    "of:00000000000000204" : {
      "segmentrouting" : {
        "name" : "s204",
        "ipv4NodeSid" : 204,
        "ipv4Loopback" : "192.168.0.204",
        "routerMac" : "00:00:00:00:02:04",
        "isEdgeRouter" : true,
        "adjacencySids" : []
      },
      "basic" : {
        "name": "s204",
        "driver" : "ofdpa-ovs"
      }
    }
  }
}
```

True for leaves. False for spines.

Device

```
{
  "devices" : {
    "of:00000000000000204" : {
      "segmentrouting" : {
        "name" : "s204",
        "ipv4NodeSid" : 204,
        "ipv4Loopback" : "192.168.0.204",
        "routerMac" : "00:00:00:00:02:04",
        "isEdgeRouter" : true,
        "adjacencySids" : []
      },
      "basic" : {
        "name" : "s204",
        "driver" : "ofdpa-ovs"
      }
    }
  }
}
```

Unused. Will be deprecated.

Device

```
{
  "devices" : {
    "of:00000000000000204" : {
      "segmentrouting" : {
        "name" : "s204",
        "ipv4NodeSid" : 204,
        "ipv4Loopback" : "192.168.0.204",
        "routerMac" : "00:00:00:00:02:04",
        "isEdgeRouter" : true,
        "adjacencySids" : []
      },
      "basic" : {
        "name": "s204",
        "driver" : "ofdpa-ovs"
      }
    }
  }
}
```

This is the name that does matter.
Will be displayed in GUI

Device

```
{
  "devices" : {
    "of:00000000000000204" : {
      "segmentrouting" : {
        "name" : "s204",
        "ipv4NodeSid" : 204,
        "ipv4Loopback" : "192.168.0.204",
        "routerMac" : "00:00:00:00:02:04",
        "isEdgeRouter" : true,
        "adjacencySids" : []
      },
      "basic" : {
        "name": "s204",
        "driver" : "ofdpa-ovs"
      }
    }
  }
}
```

Driver.

ofdpa-ovs for OpenvSwitch.

ofdpa3 for Broadcom XGS (Trident2, Tomahawk)

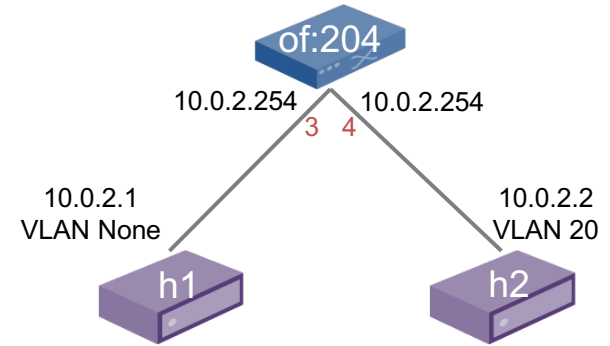
qmx-ofdpa3 for Broadcom QMX

Port - VLAN Untagged

```
{  
  "ports" : {  
    "of:00000000000000204/3" : {  
      "interfaces" : [  
        {  
          "ips" : ["10.0.2.254/24"],  
          "vlan-untagged": 20  
        }  
      ]  
    }  
  }  
}
```

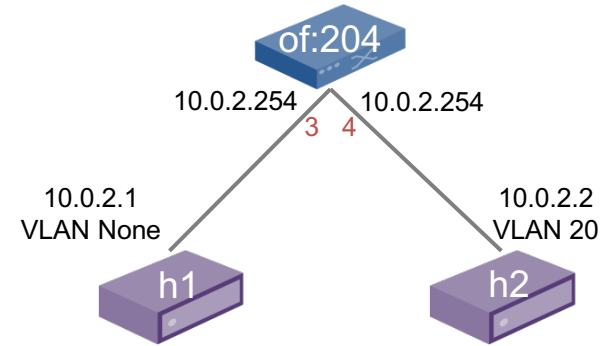
DPID/port number

Equivalent to access port.
Expecting untagged packet and will push VLAN 20



Port - VLAN Tagged

```
{  
  "ports" : {  
    "of:00000000000000204/4" : {  
      "interfaces" : [  
        {  
          "ips" : ["10.0.2.254/24"],  
          "vlan-tagged": [20]  
        }  
      ]  
    }  
  }  
}
```



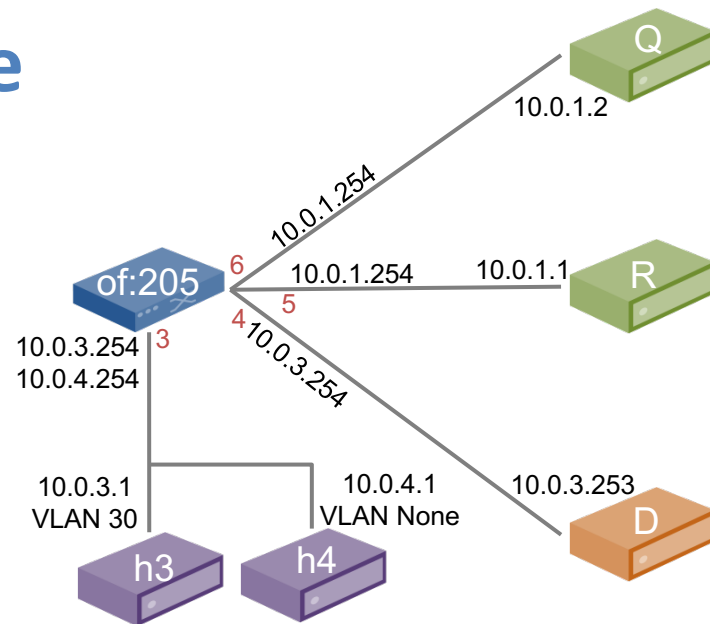
Equivalent to trunk port.

Expecting VLAN 20 tagged packet and will keep the tag.

It is possible to have more than one trunk VLAN.

Port - VLAN Tagged with Native

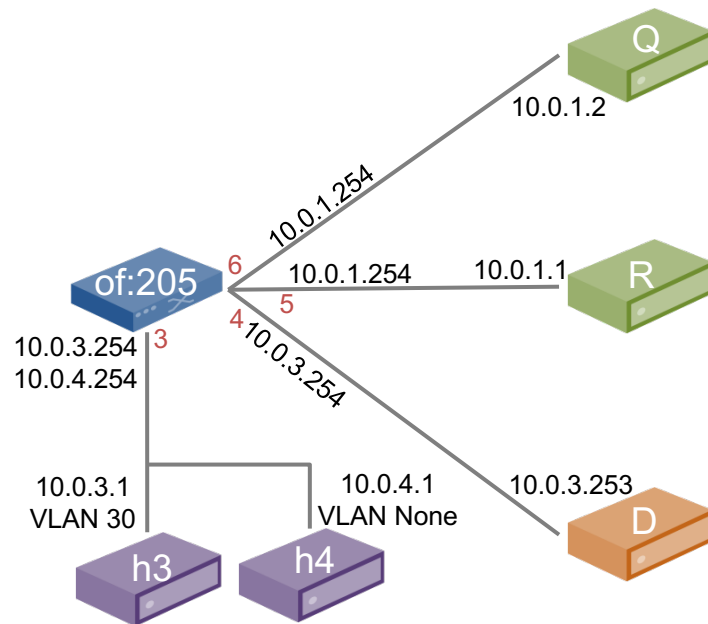
```
{  
  "ports" : {  
    "of:00000000000000205/3" : {  
      "interfaces" : [  
        {  
          "ips" : ["10.0.3.254/24", "10.0.4.254/24"]  
          "vlan-tagged" : [30],  
          "vlan-native" : 40  
        }  
      ]  
    }  
  }  
}
```



Equivalent to trunk port with native VLAN.
Expecting VLAN 30 tagged packet and will keep the tag.
Also expecting untagged packet and will push VLAN 40

Port - Subnet

```
{
  "ports" : {
    "of:00000000000000205/3" : {
      "interfaces" : [
        {
          "ips" : ["10.0.3.254/24", "10.0.4.254/24"]
          "vlan-tagged" : [30],
          "vlan-native" : 40
        }
      ]
    }
  }
}
```



Two information can be derived from the subnet config:

- (1) The interface IP on the switch is 10.0.3.254
- (2) The subnet on this interface is 10.0.3.0/24

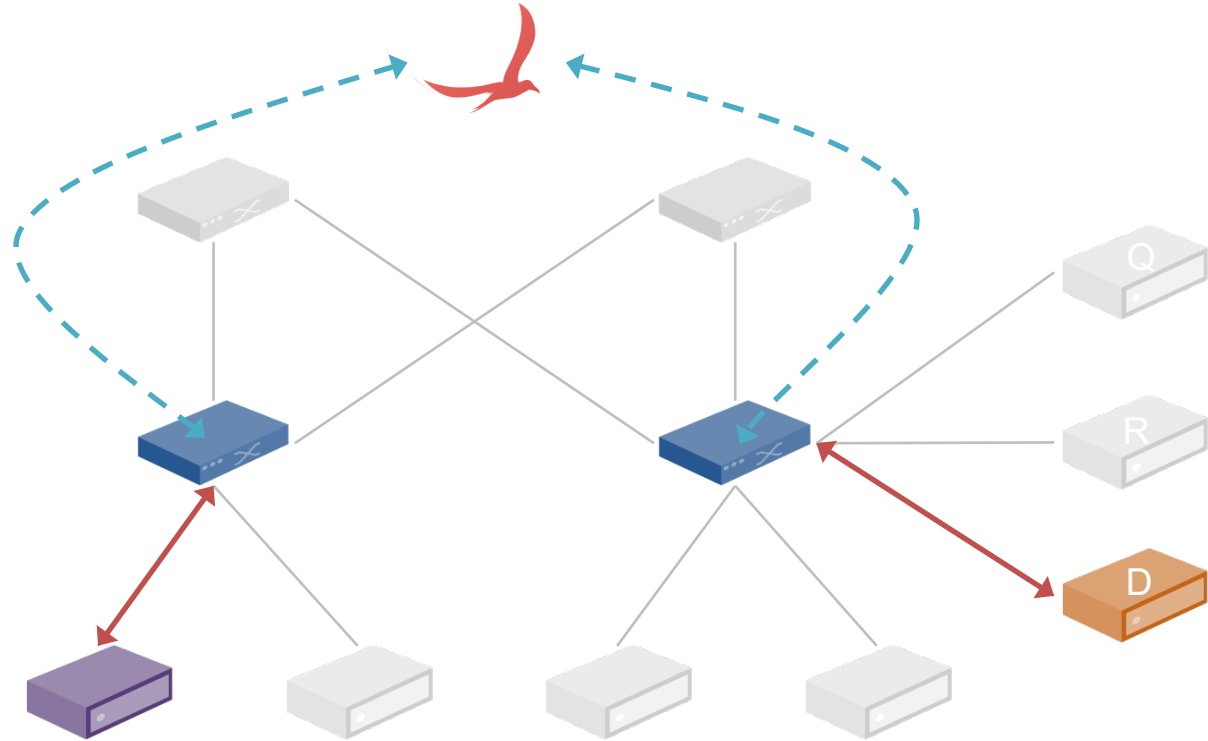
Currently we don't map between subnet and VLAN.

VLAN information is solely used for defining bridging domain.

Subnet information is solely used for routing.

There are some confusions. Will have 1-to-1 mapping in the future.

L3 DHCP Relay



L3 DHCP Relay

```
{
  "apps" : {
    "org.onosproject.dhcprelay" : {
      "default": [
        {
          "dhcpServerConnectPoint": "of:0000000000000205/6",
          "serverIps": ["10.0.3.253"]
        }
      ]
    }
  }
}
```

The location where the DHCP server is attached to

L3 DHCP Relay

```
{  
  "apps" : {  
    "org.onosproject.dhcprelay" : {  
      "default": [  
        {  
          "dhcpServerConnectPoint": "of:00000000000000205/6",  
          "serverIps": ["10.0.3.253"]  
        }  
      ]  
    }  
  }  
}
```

The IP address of DHCP server

dhcpcd.conf

```
subnet 10.0.2.0 netmask 255.255.255.0 {  
    range 10.0.2.100 10.0.2.240;  
    option routers 10.0.2.254;  
}
```

```
subnet 10.0.3.0 netmask 255.255.255.0 {  
    range 10.0.3.100 10.0.3.240;  
    option routers 10.0.3.254;  
}
```

Subnets we want to assign

```
subnet 10.0.4.0 netmask 255.255.255.0 {  
    range 10.0.4.100 10.0.4.240;  
    option routers 10.0.4.254;  
}
```

```
host h1 {  
    hardware ethernet 00:aa:00:00:00:01;  
    fixed-address 10.0.2.1;  
}
```

dhcpcd.conf

```
subnet 10.0.2.0 netmask 255.255.255.0 {  
  range 10.0.2.100 10.0.2.240;  
  option routers 10.0.2.254;  
}
```

```
subnet 10.0.3.0 netmask 255.255.255.0 {  
  range 10.0.3.100 10.0.3.240;  
  option routers 10.0.3.254;  
}
```

```
subnet 10.0.4.0 netmask 255.255.255.0 {  
  range 10.0.4.100 10.0.4.240;  
  option routers 10.0.4.254;  
}
```

```
host h1 {  
  hardware ethernet 00:aa:00:00:00:01;  
  fixed-address 10.0.2.1;  
}
```

IP range that will be used for this subnet

dhcpcd.conf

```
subnet 10.0.2.0 netmask 255.255.255.0 {  
    range 10.0.2.100 10.0.2.240;  
    option routers 10.0.2.254;  
}
```

```
subnet 10.0.3.0 netmask 255.255.255.0 {  
    range 10.0.3.100 10.0.3.240;  
    option routers 10.0.3.254;  
}
```

```
subnet 10.0.4.0 netmask 255.255.255.0 {  
    range 10.0.4.100 10.0.4.240;  
    option routers 10.0.4.254;  
}
```

```
host h1 {  
    hardware ethernet 00:aa:00:00:00:01;  
    fixed-address 10.0.2.1;  
}
```

Default gateway we want to assign, which is the interface IP of the leaf switch

dhcpcd.conf

```
subnet 10.0.2.0 netmask 255.255.255.0 {  
    range 10.0.2.100 10.0.2.240;  
    option routers 10.0.2.254;  
}
```

```
subnet 10.0.3.0 netmask 255.255.255.0 {  
    range 10.0.3.100 10.0.3.240;  
    option routers 10.0.3.254;  
}
```

```
subnet 10.0.4.0 netmask 255.255.255.0 {  
    range 10.0.4.100 10.0.4.240;  
    option routers 10.0.4.254;  
}
```

```
host h1 {  
    hardware ethernet 00:aa:00:00:00:01;  
    fixed-address 10.0.2.1;  
}
```

We can also have static IP assignment

vRouter

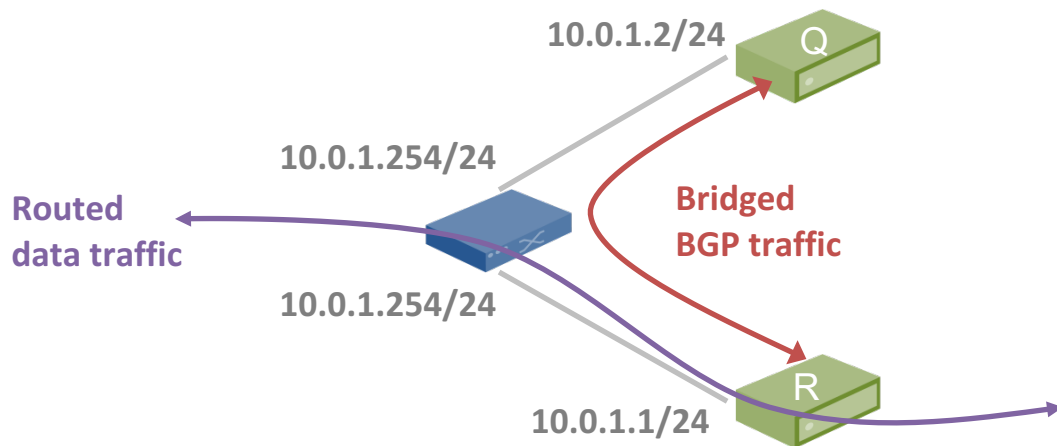
```
{
  "ports" : {
    "of:00000000000000205/6" : {
      "interfaces" : [
        {
          "ips" : ["10.0.1.254/24"]
          "vlan-untagged": 10
        }
      ]
    },
    "of:00000000000000205/7" : {
      "interfaces" : [
        {
          "ips" : ["10.0.1.254/24"]
          "vlan-untagged": 10
        }
      ]
    }
  }
}
```

Quagga and upstream router need to be in the same VLAN and subnet on the same leaf.

We need 3 IPs in the subnet. 1 for Quagga, 1 for upstream router, 1 for leaf switch interface.
/29 is the minimum subnet.

Route Map

- Quagga announces **leaf switch**, instead of quagga itself, as the **next hop** to the internal network



Quagga - bgpd.conf

```
ip prefix-list 1 seq 10 permit 10.0.2.0/24
ip prefix-list 1 seq 20 permit 10.0.3.0/24
```

List of prefixes we want to announce

```
route-map NEXTHOP41 permit 10
match ip address prefix-list 1
set ip next-hop 10.0.1.254
```

```
neighbor 10.0.1.1
remote-as 65001
neighbor 10.0.1.1 ebgp-multihop
neighbor 10.0.1.1 timers connect 5
neighbor 10.0.1.1 advertisement-interval 5
neighbor 10.0.1.1 route-map NEXTHOP41 out
```

```
network 10.0.2.0/24
network 10.0.3.0/24
```

Quagga - bgpd.conf

```
ip prefix-list 1 seq 10 permit 10.0.2.0/24
ip prefix-list 1 seq 20 permit 10.0.3.0/24
```

```
route-map NEXTHOP41 permit 10
match ip address prefix-list 1
set ip next-hop 10.0.1.254
```

Define the route map and its name

```
neighbor 10.0.1.1
remote-as 65001
neighbor 10.0.1.1 ebgp-multihop
neighbor 10.0.1.1 timers connect 5
neighbor 10.0.1.1 advertisement-interval 5
neighbor 10.0.1.1 route-map NEXTHOP41 out
```

```
network 10.0.2.0/24
network 10.0.3.0/24
```

Quagga - bgpd.conf

```
ip prefix-list 1 seq 10 permit 10.0.2.0/24
ip prefix-list 1 seq 20 permit 10.0.3.0/24
```

```
route-map NEXTHOP41 permit 10
match ip address prefix-list 1
set ip next-hop 10.0.1.254
```

Apply the route map to the prefixes in prefix-list 1

```
neighbor 10.0.1.1
remote-as 65001
neighbor 10.0.1.1 ebgp-multihop
neighbor 10.0.1.1 timers connect 5
neighbor 10.0.1.1 advertisement-interval 5
neighbor 10.0.1.1 route-map NEXTHOP41 out
```

```
network 10.0.2.0/24
network 10.0.3.0/24
```

Quagga - bgpd.conf

```
ip prefix-list 1 seq 10 permit 10.0.2.0/24
ip prefix-list 1 seq 20 permit 10.0.3.0/24
```

```
route-map NEXTHOP41 permit 10
match ip address prefix-list 1
set ip next-hop 10.0.1.254
```

Set the leaf switch, instead of Quagga itself, as the next hop for these prefixes

```
neighbor 10.0.1.1
remote-as 65001
neighbor 10.0.1.1 ebgp-multihop
neighbor 10.0.1.1 timers connect 5
neighbor 10.0.1.1 advertisement-interval 5
neighbor 10.0.1.1 route-map NEXTHOP41 out
```

```
network 10.0.2.0/24
network 10.0.3.0/24
```

Quagga - bgpd.conf

```
ip prefix-list 1 seq 10 permit 10.0.2.0/24
ip prefix-list 1 seq 20 permit 10.0.3.0/24
```

```
route-map NEXTHOP41 permit 10
match ip address prefix-list 1
set ip next-hop 10.0.1.254
```

```
neighbor 10.0.1.1
remote-as 65001
neighbor 10.0.1.1 ebgp-multihop
neighbor 10.0.1.1 timers connect 5
neighbor 10.0.1.1 advertisement-interval 5
neighbor 10.0.1.1 route-map NEXTHOP41 out
```

```
network 10.0.2.0/24
network 10.0.3.0/24
```

BGP peer information

Quagga - bgpd.conf

```
ip prefix-list 1 seq 10 permit 10.0.2.0/24
ip prefix-list 1 seq 20 permit 10.0.3.0/24
```

```
route-map NEXTHOP41 permit 10
match ip address prefix-list 1
set ip next-hop 10.0.1.254
```

```
neighbor 10.0.1.1
remote-as 65001
neighbor 10.0.1.1 ebgp-multihop
neighbor 10.0.1.1 timers connect 5
neighbor 10.0.1.1 advertisement-interval 5
neighbor 10.0.1.1 route-map NEXTHOP41 out
```

Announce the route map to the neighbor

```
network 10.0.2.0/24
network 10.0.3.0/24
```

Quagga - bgpd.conf

```
ip prefix-list 1 seq 10 permit 10.0.2.0/24
ip prefix-list 1 seq 20 permit 10.0.3.0/24
```

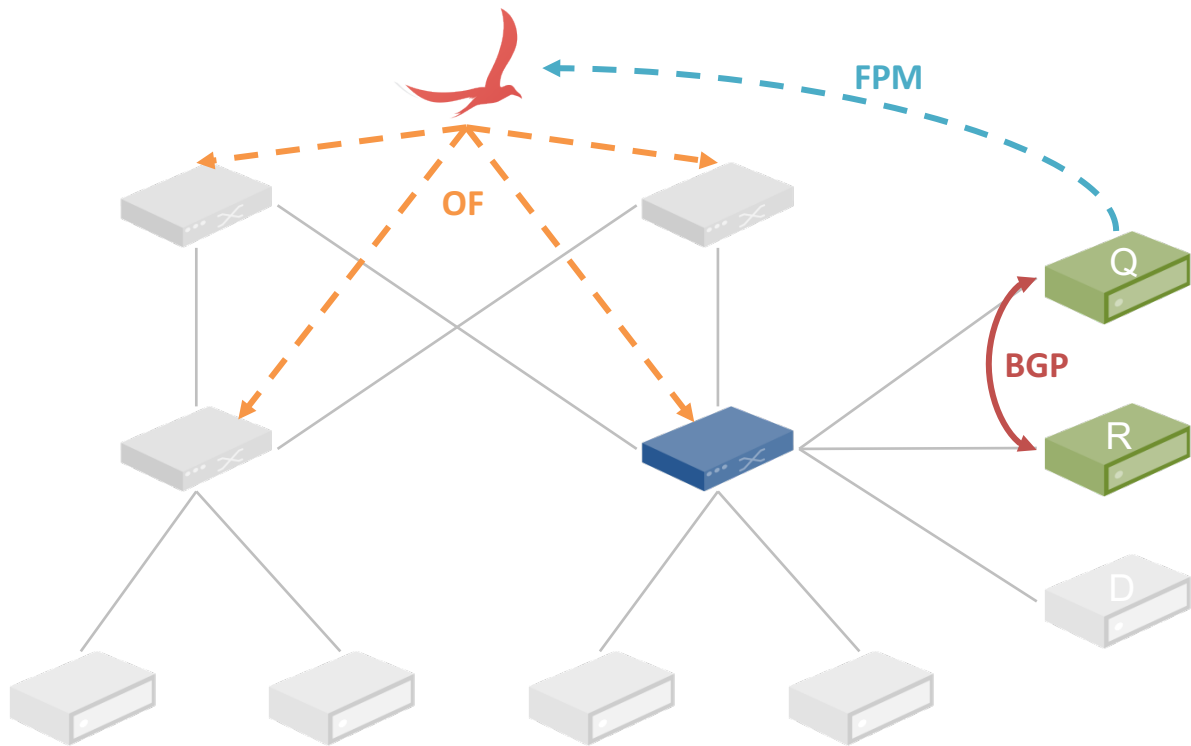
```
route-map NEXTHOP41 permit 10
match ip address prefix-list 1
set ip next-hop 10.0.1.254
```

```
neighbor 10.0.1.1
remote-as 65001
neighbor 10.0.1.1 ebgp-multihop
neighbor 10.0.1.1 timers connect 5
neighbor 10.0.1.1 advertisement-interval 5
neighbor 10.0.1.1 route-map NEXTHOP41 out
```

```
network 10.0.2.0/24
network 10.0.3.0/24
```

Networks that are reachable from Quagga

FPM (Forwarding Plane Manager)



Quagga - zebrad.conf

```
ip route 0.0.0.0/0 172.16.0.1
```

(Optional) Point default route to management network

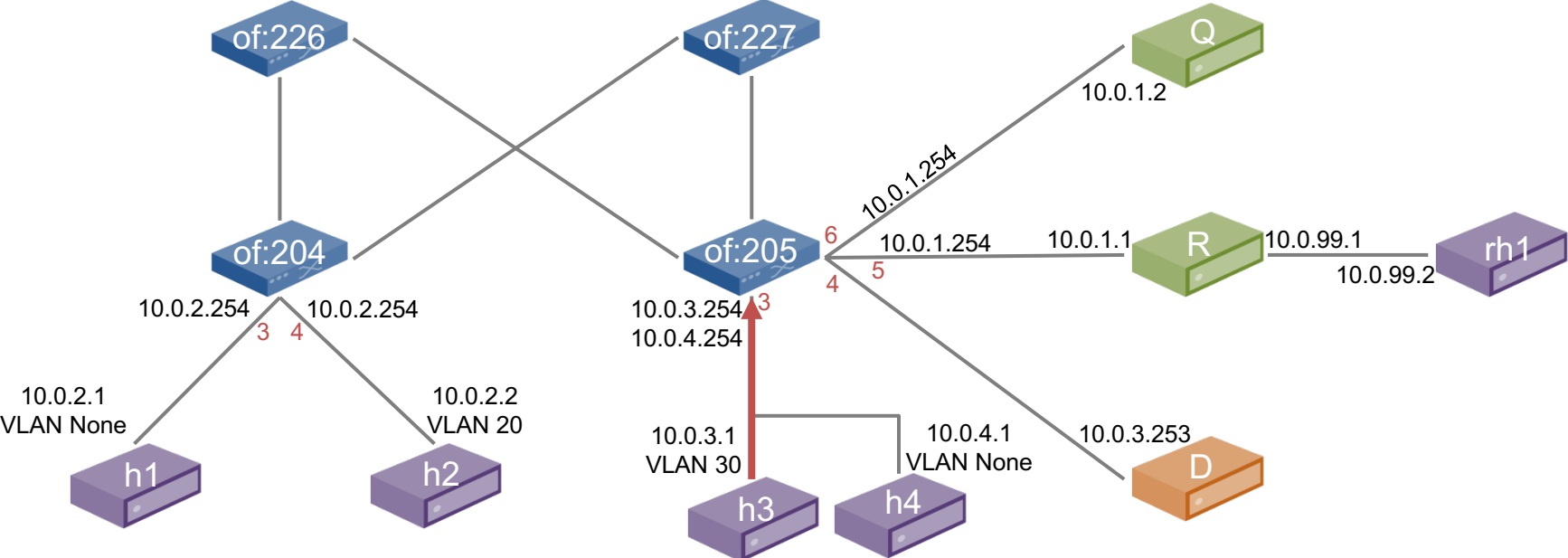
```
fpm connection ip 192.168.56.11 port 2620
```

FPM connection to ONOS IP.

Need to be reachable from Quagga

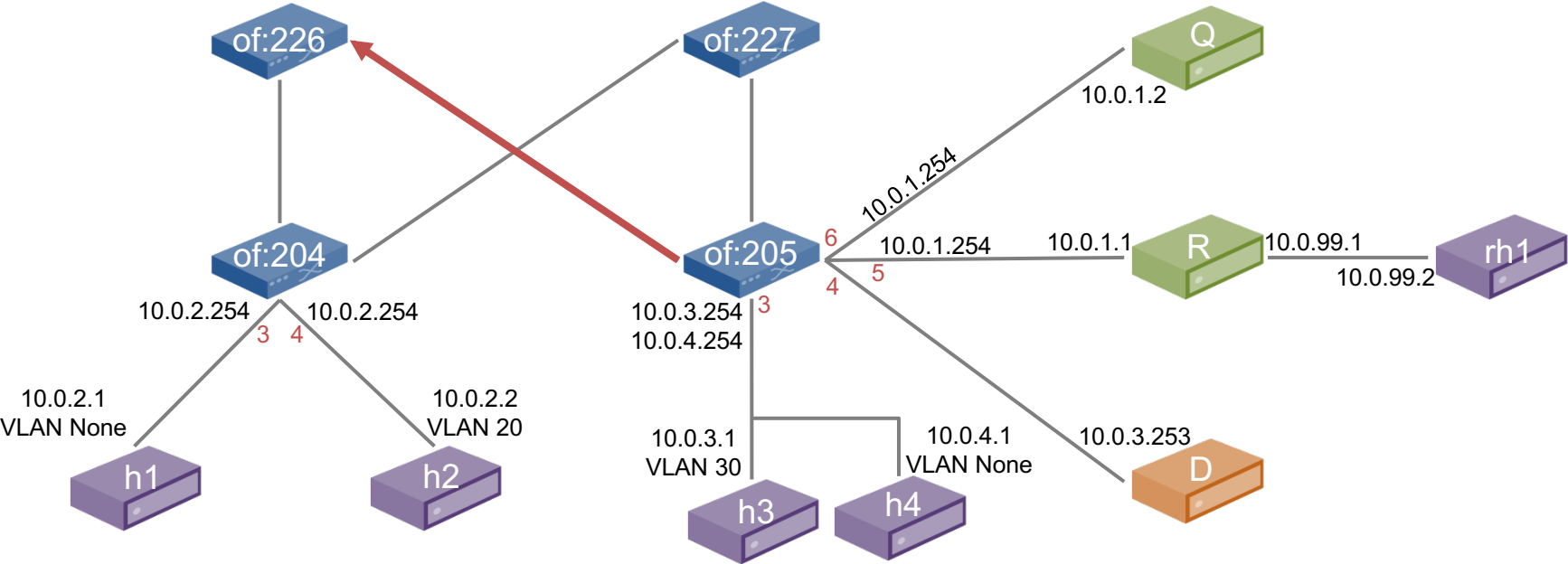
Demo Topology

VLAN: 30
Src mac: h3
Dst mac: routerMac of of:205
Src IP: h3
Dst IP: h1



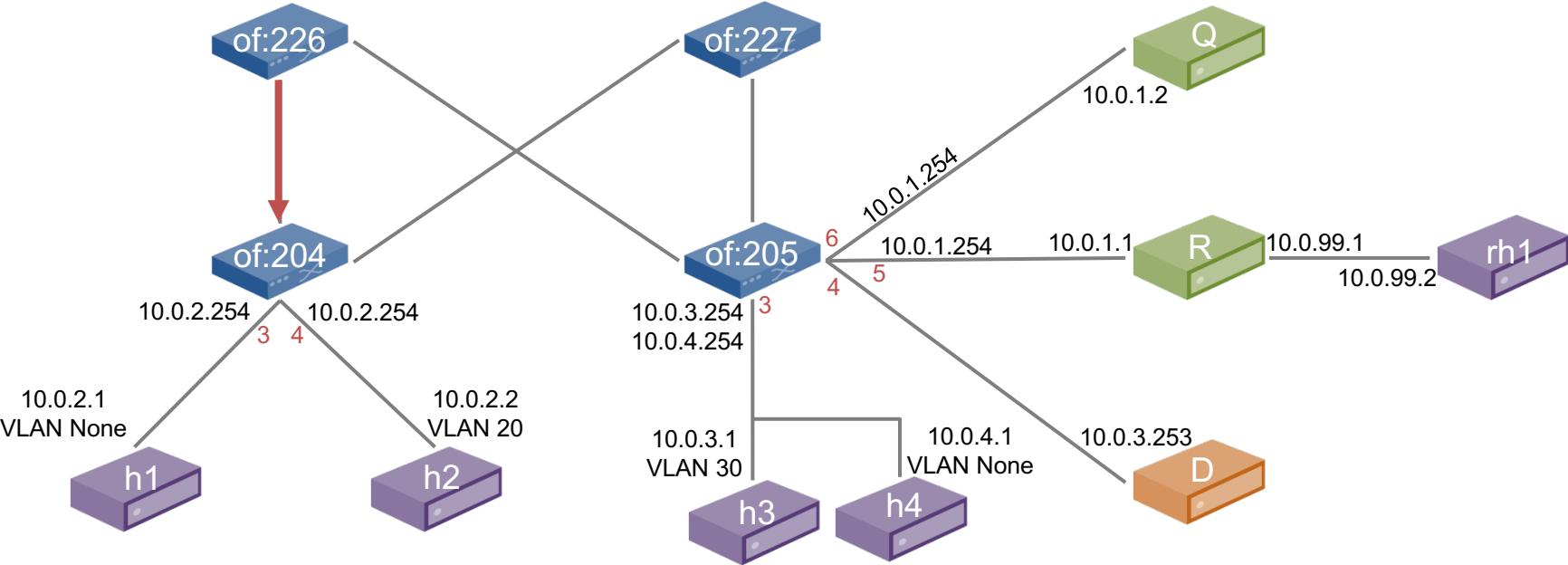
Demo Topology

Src mac: routerMac of of:205
Dst mac: routerMac of of:226
MPLS label: 204
Src IP: h3
Dst IP: h1



Demo Topology

Src mac: routerMac of of:226
Dst mac: routerMac of of:204
Src IP: h3
Dst IP: h1

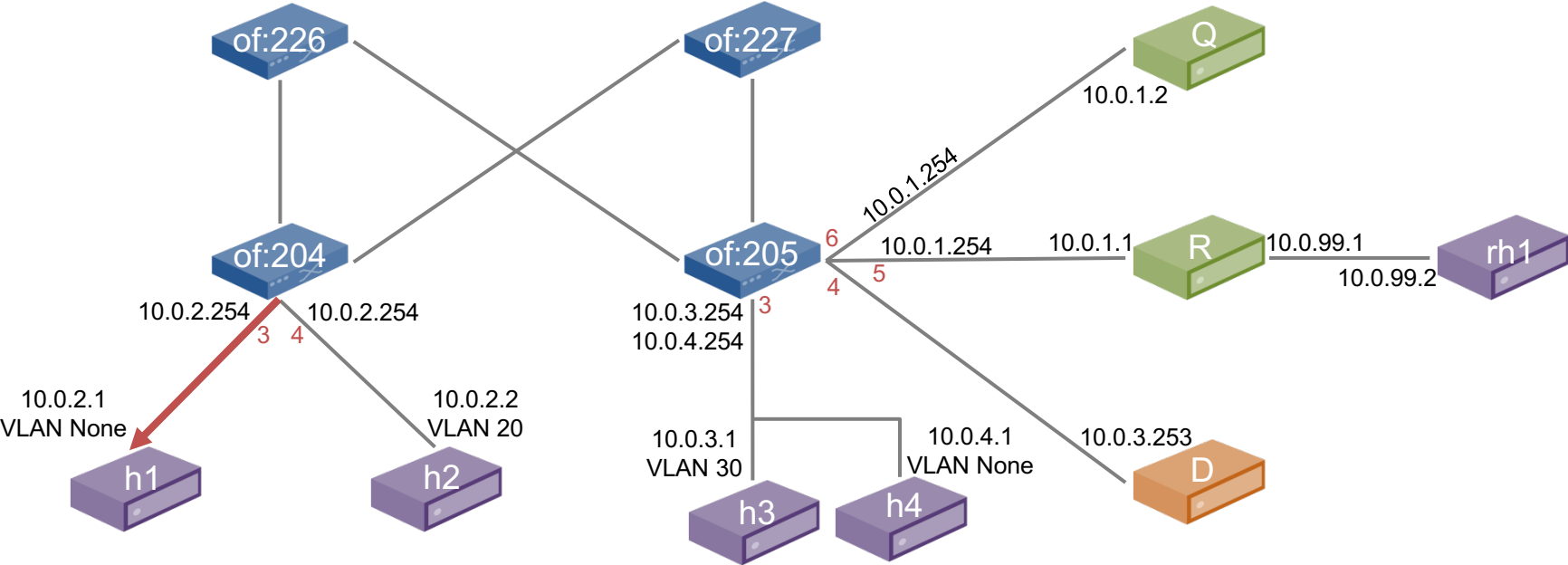


Q: Quagga, R: Upstream Router, D: DHCP Server



Demo Topology

Src mac: routerMac of of:204
Dst mac: h1
Src IP: h3
Dst IP: h1



Q: Quagga, R: Upstream Router, D: DHCP Server



Demo

Scripts and configurations will be available at

<https://github.com/opennetworkinglab/routing/tree/master/trellis>
after this event

Thank you !

Visit <http://wiki.opencord.org> for more information