Kubernetes Netowrking

AbdulKarim - ReLambda
[relambda.com]

v1.0.2

Introduction

Agenda

- Kubernetes - brief intro
- Pods
  - Communication Using Pod IP
  - Multi Node
  - Overlay Networks
- Services
  - Communication Using Service IP
  - IPTables
- QnA

K8s is

Open Source

Container Orchestrator

AutoScaling

Self Healing

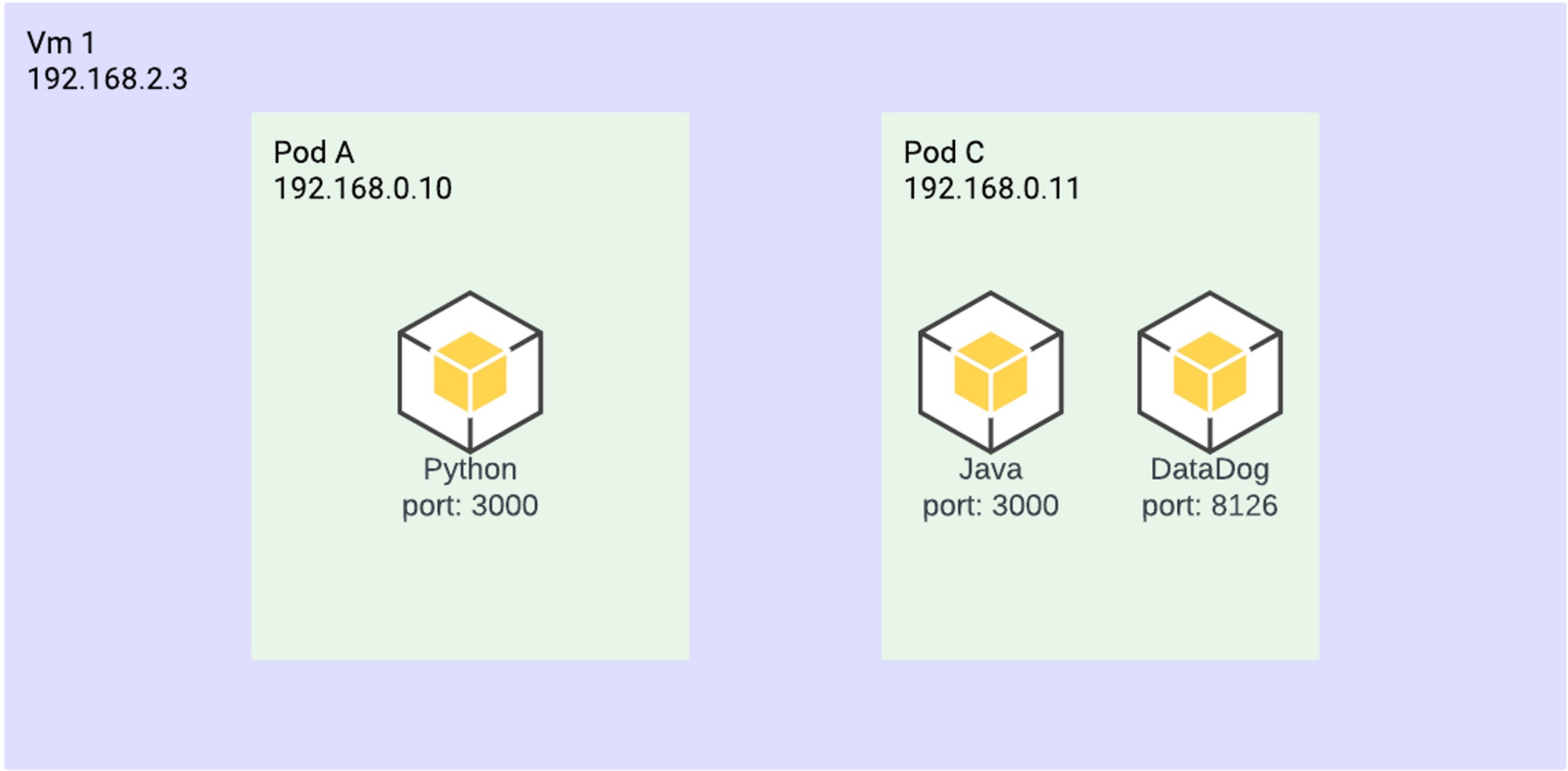
- There is no such thing as a **container**
- **Pod** is a group of one/many **containers**
  - Guarded by **CGroups**
  - Isolated by **Namespaces**
- Pods are **Atomic**
- Pods are **Ephemeral**

Every Pod

Gets Its

Own Unique IP

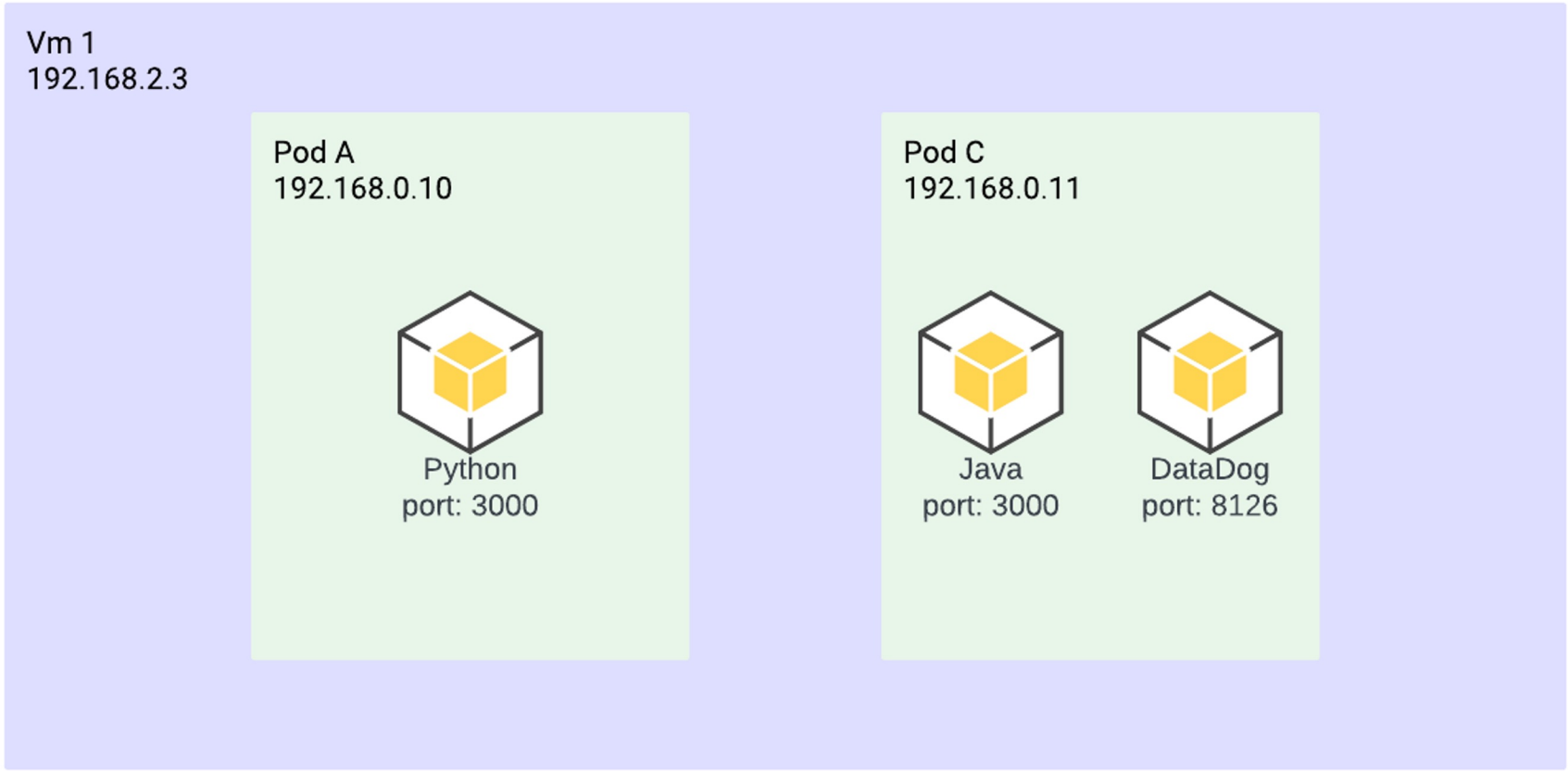
(IP-Per-Pod)


Vm 1
192.168.2.3
Pod A
192.168.0.10
Python
port: 3000
Pod C
192.168.0.11
Java
port: 3000
DataDog
port: 8126

All **Containers**

In **A Pod** Share The

Same Network Namespace


Vm 1
192.168.2.3
Pod A
192.168.0.10
Python
port: 3000
Pod C
192.168.0.11
Java
port: 3000
DataDog
port: 8126

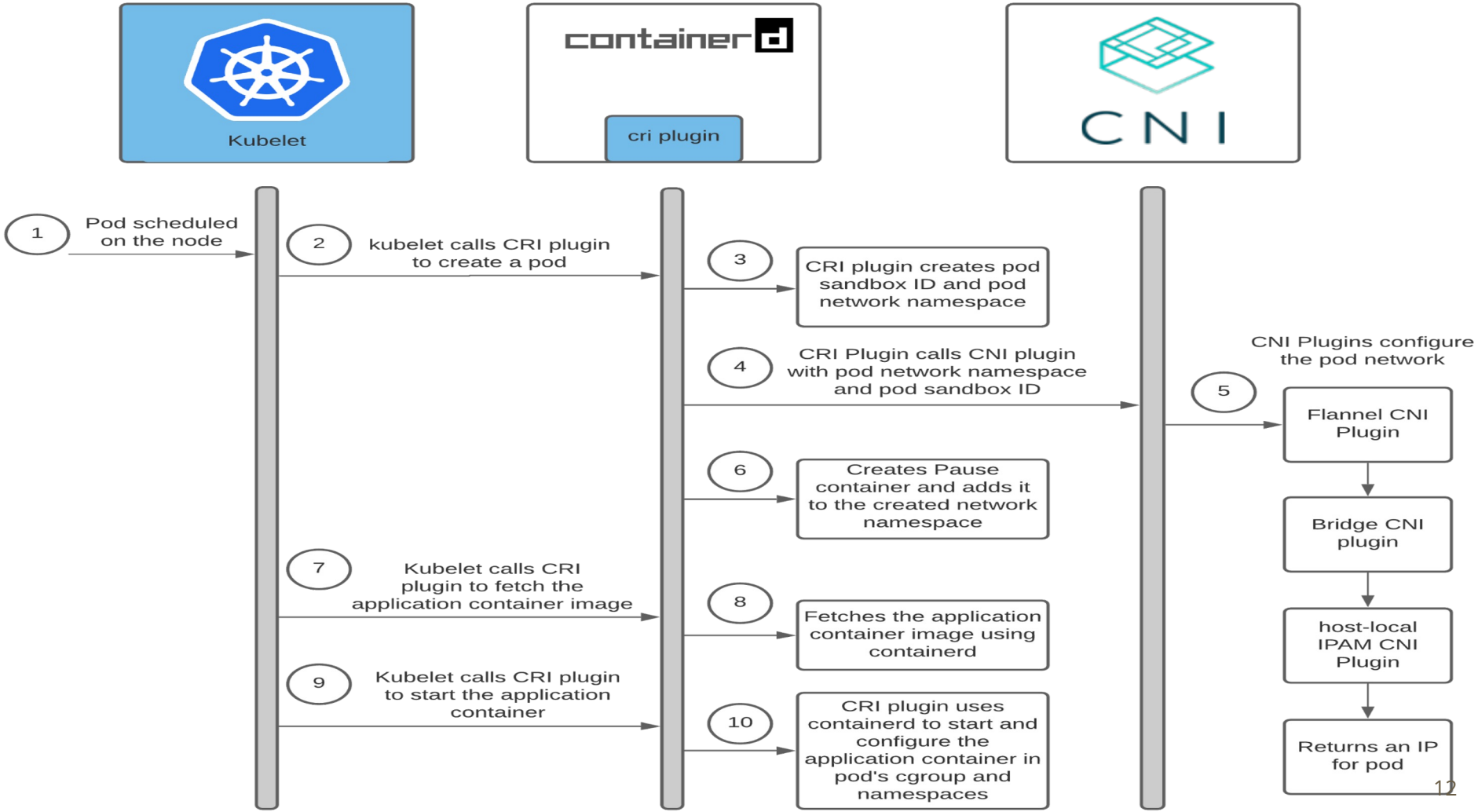
Independant Address Spaces

Host IP Range

Pod IP Range

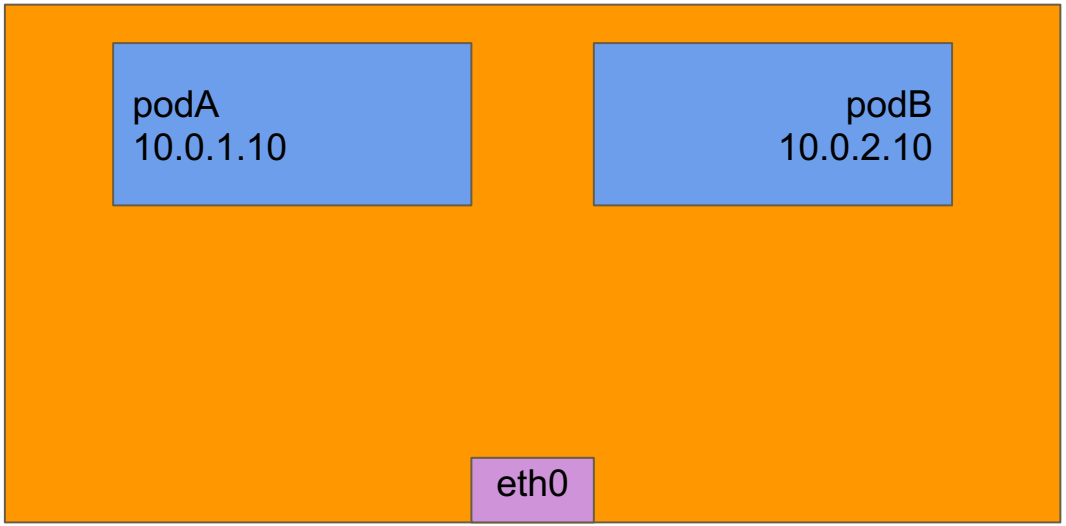
Service IP Range

How Does A Pod Get An IP Address?

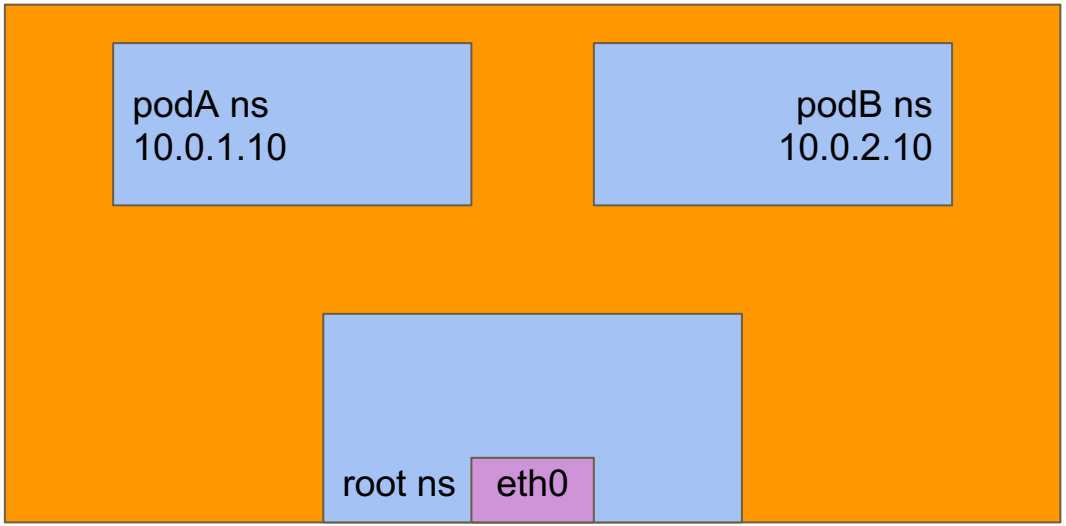


https://ronaknathani.com/blog/2020/08/how-a-kubernetes-pod-gets-an-ip-address/

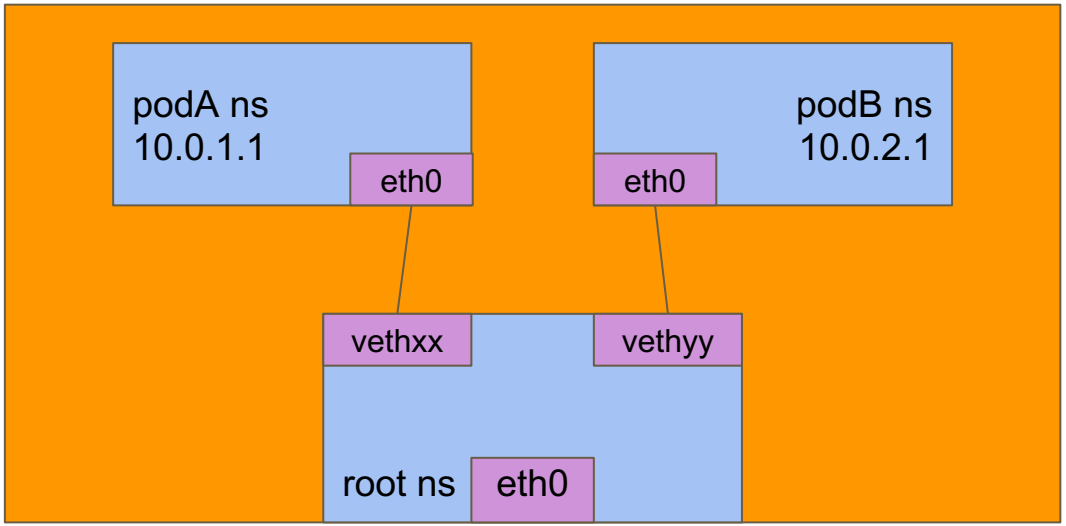
How Does A Pod Talk To Another Pod using Pod-IP?



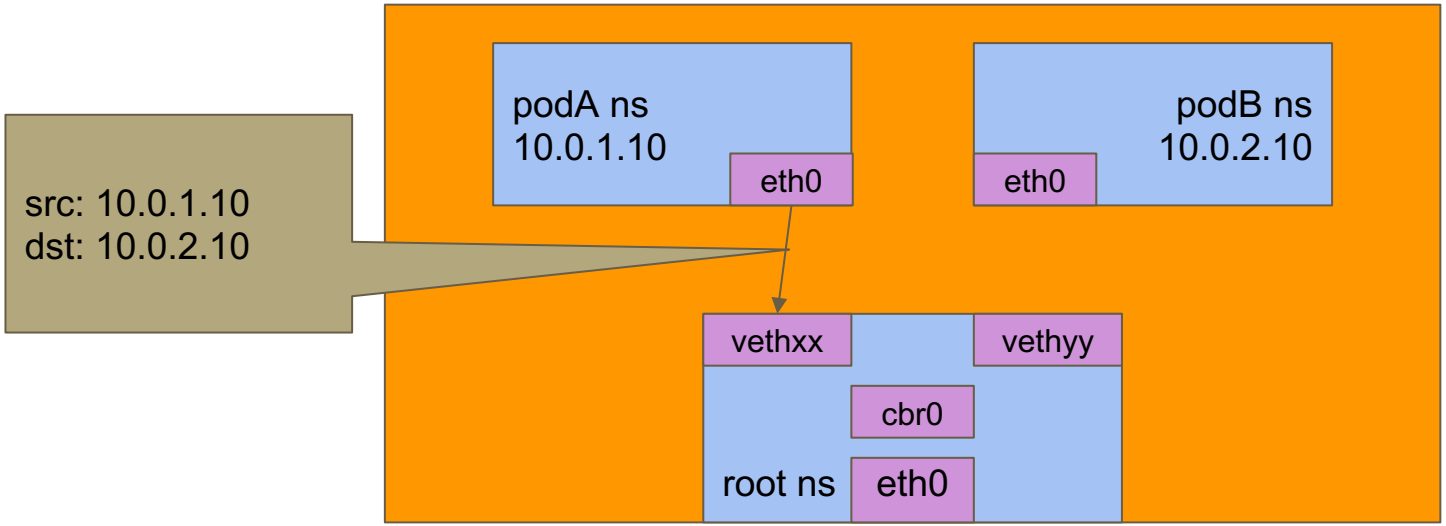
Node1 (10.0.240.10)



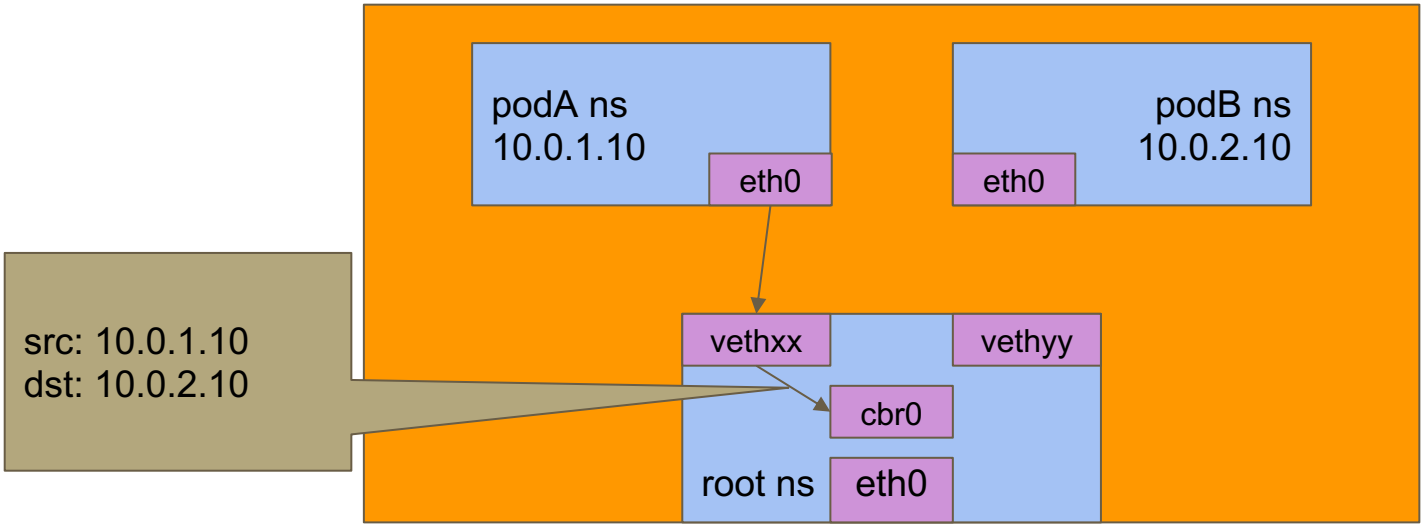
Node1 (10.0.240.10)



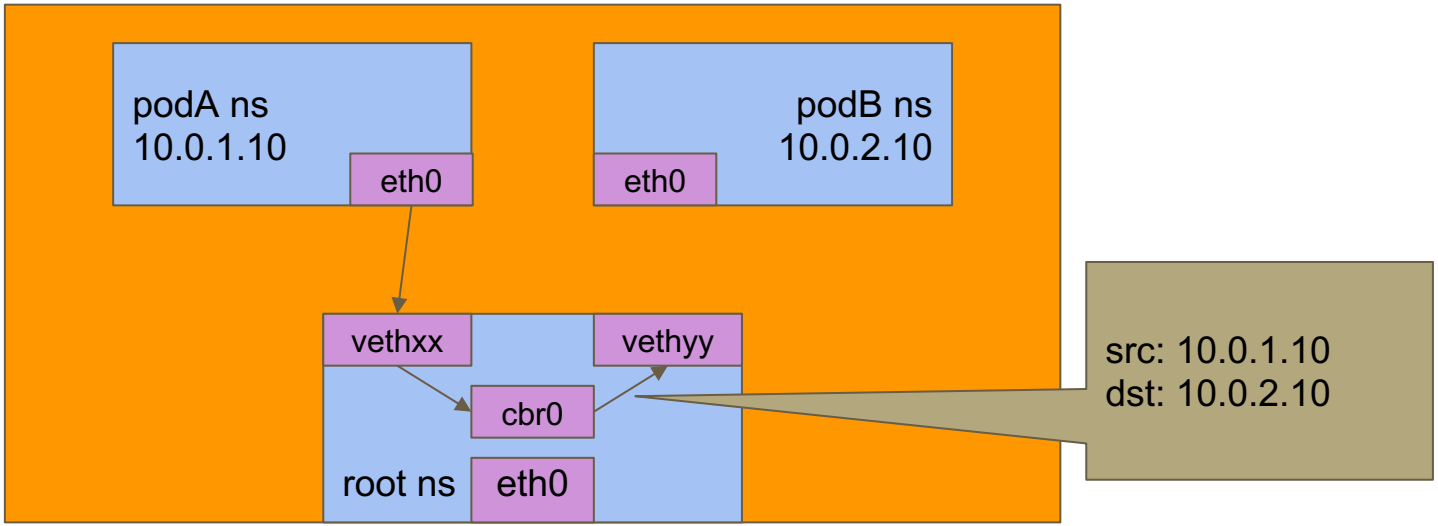
Node1 (10.0.240.1)



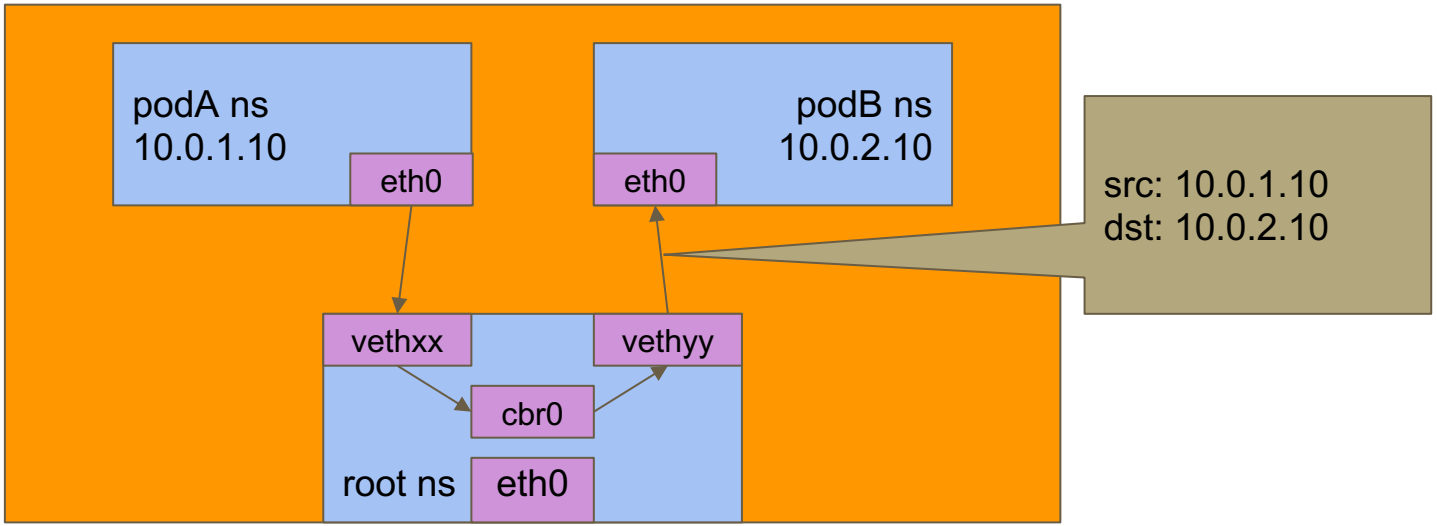
Node1 (10.0.240.10)



Node1 (10.0.240.10)

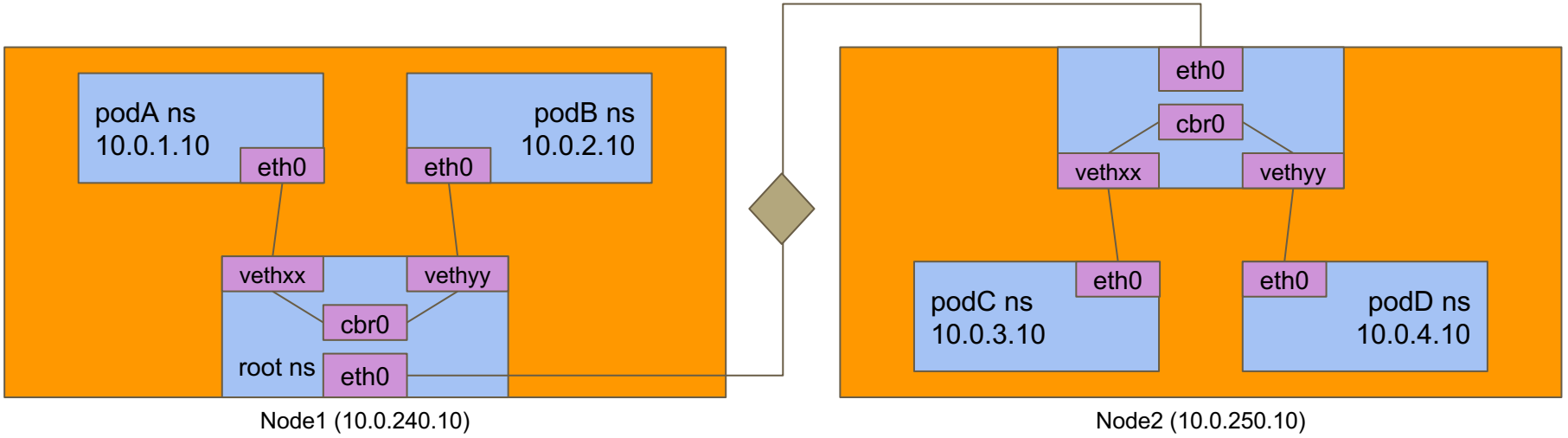


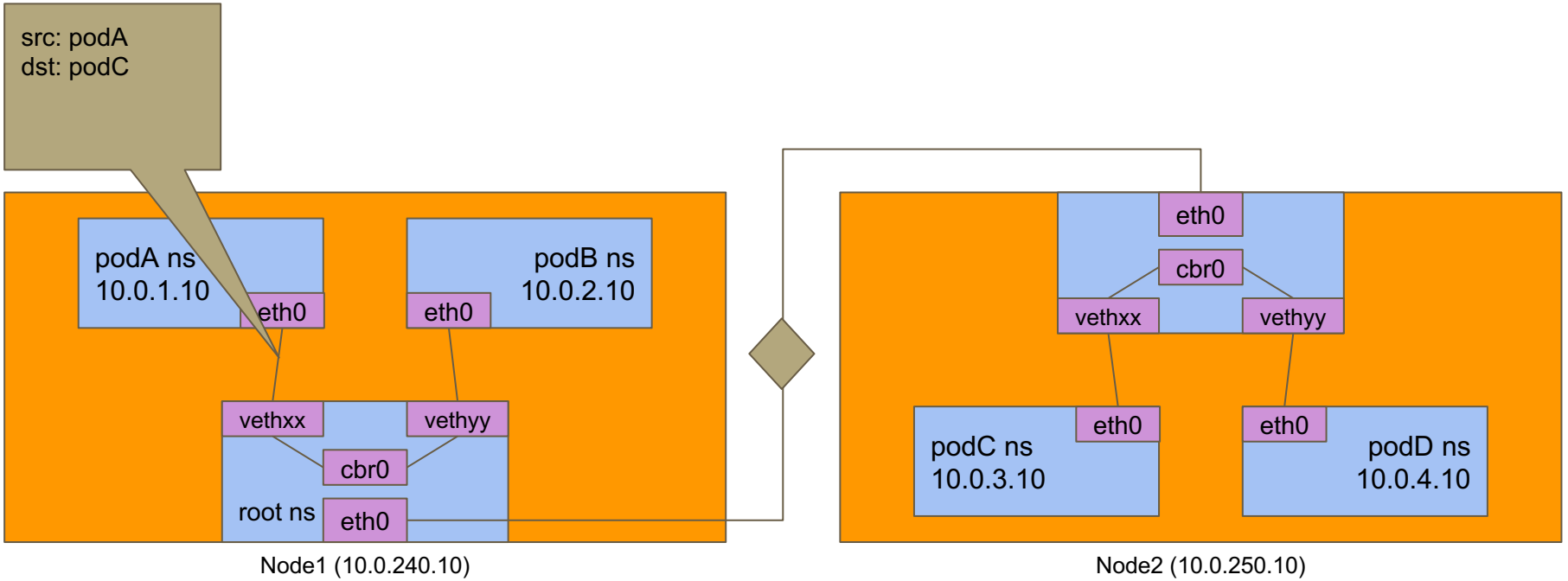
Node1 (10.0.240.10)

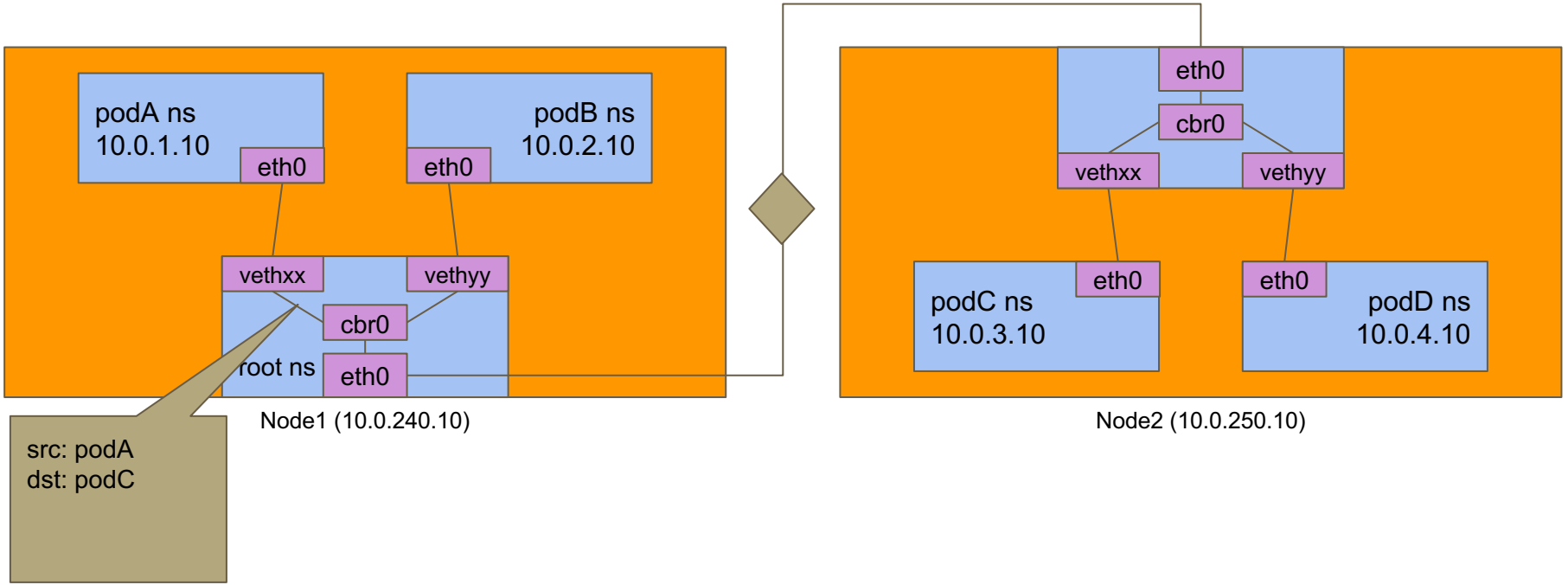


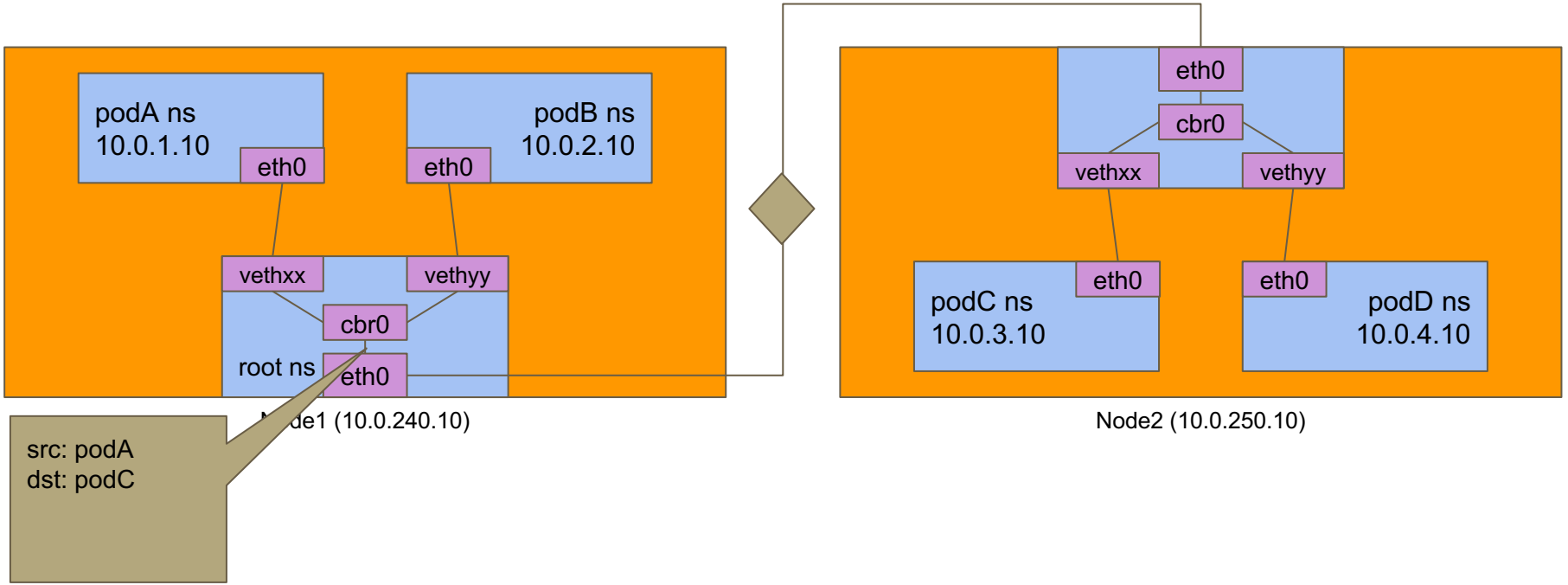
Node1 (10.0.240.10)

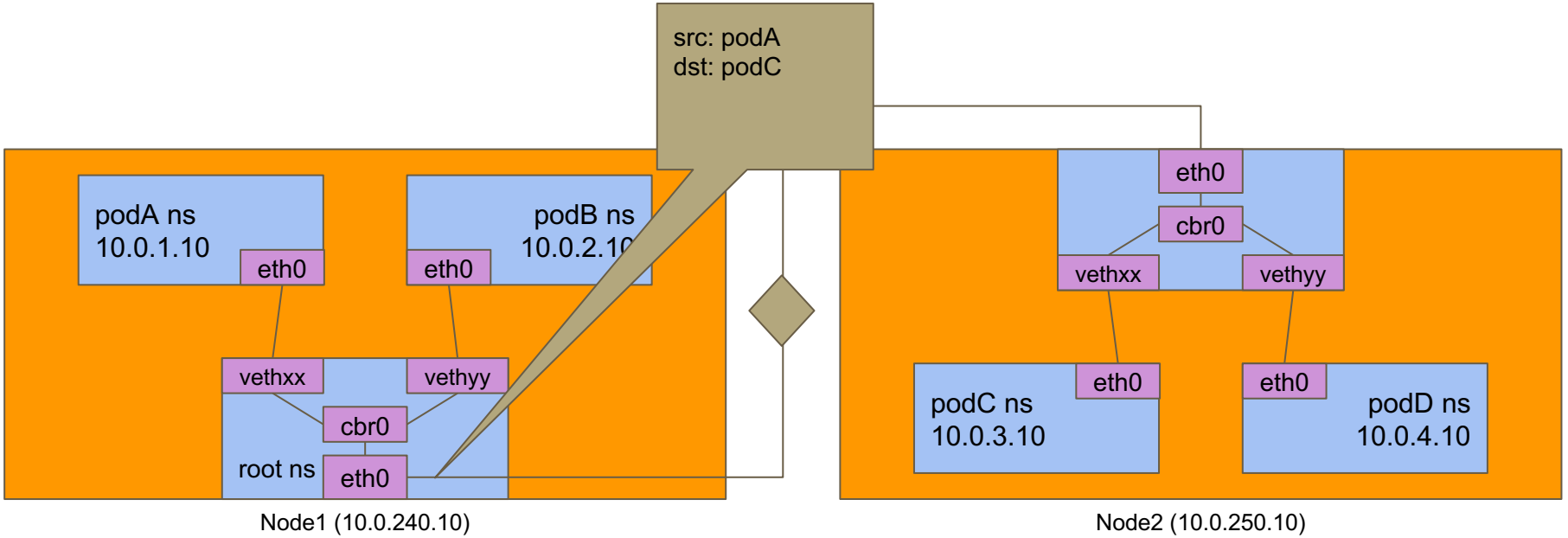
How Does A Pod Talk To Another Pod using Pod-IP?

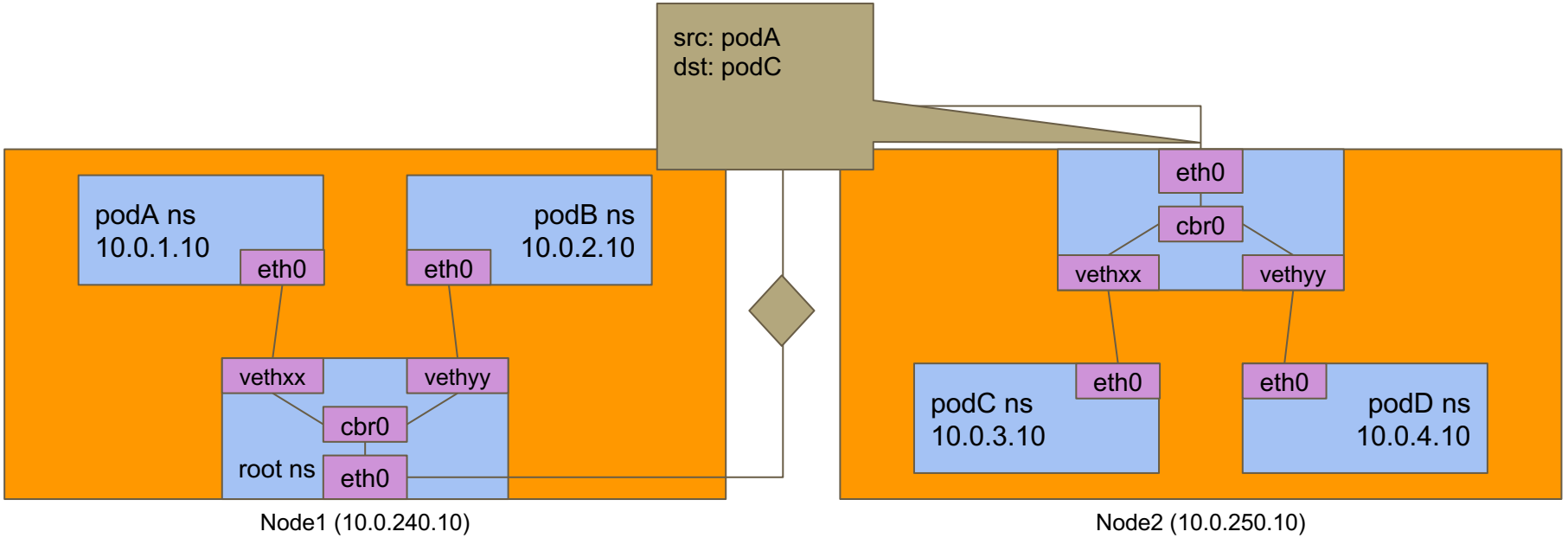

podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
root ns
eth0
Node1 (10.0.240.10)
eth0
cbr0
vethxx
vethyy
podC ns
10.0.3.10
eth0
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)


src: podA
dst: podC
podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
root ns
eth0
Node1 (10.0.240.10)
eth0
cbr0
vethxx
vethyy
podC ns
10.0.3.10
eth0
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)


podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
root ns
eth0
Node1 (10.0.240.10)
src: podA
dst: podC
eth0
cbr0
vethxx
vethyy
podC ns
10.0.3.10
eth0
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)


podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
root ns
eth0
de1 (10.0.240.10)
src: podA
dst: podC
eth0
cbr0
vethxx
vethyy
podC ns
10.0.3.10
eth0
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)


src: podA
dst: podC
podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
root ns
eth0
Node1 (10.0.240.10)
eth0
cbr0
vethxx
vethyy
podC ns
10.0.3.10
eth0
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)


src: podA
dst: podC
podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
root ns
eth0
Node1 (10.0.240.10)
eth0
cbr0
vethxx
vethyy
podC ns
10.0.3.10
eth0
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)

Name	Instance ID	Instance state	Instance type	Private IP address	Security group name	Launch time	VPC ID
[redacted]node-group-1	i-[redacted]0bcec0970	Running	m5.2xlarge	192.150.2.21	[redacted]eks-2LgdDngQ-node-20230611092714934500000008	2023/07/13 19:58 GMT+5:30	vpc-0fae52340a

Instance: i-041c01e00bcec0970 (sss-prod-node-group-1)

Public IPv4 address
–

Public IPv4 DNS
–

Subnet ID
subnet-03538cbee0b82718a ([redacted]eks-vpc-private-ap-south-1b)

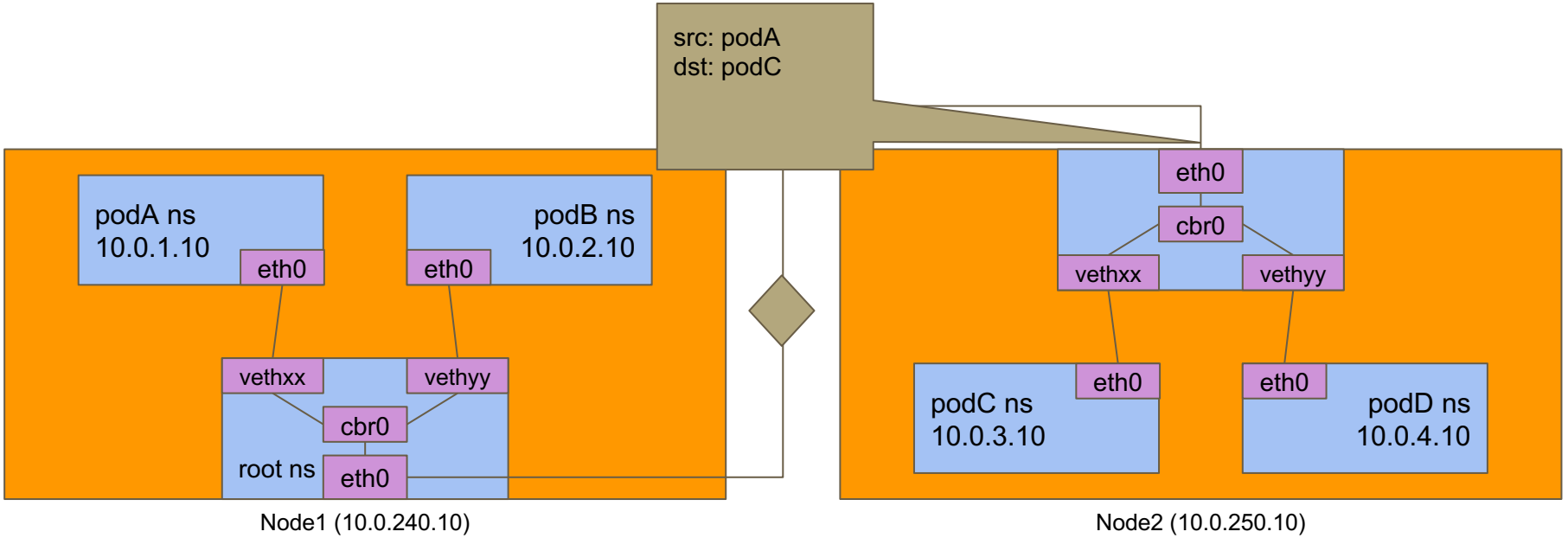
Private IPv4 addresses
192.150.2.21
192.150.2.187

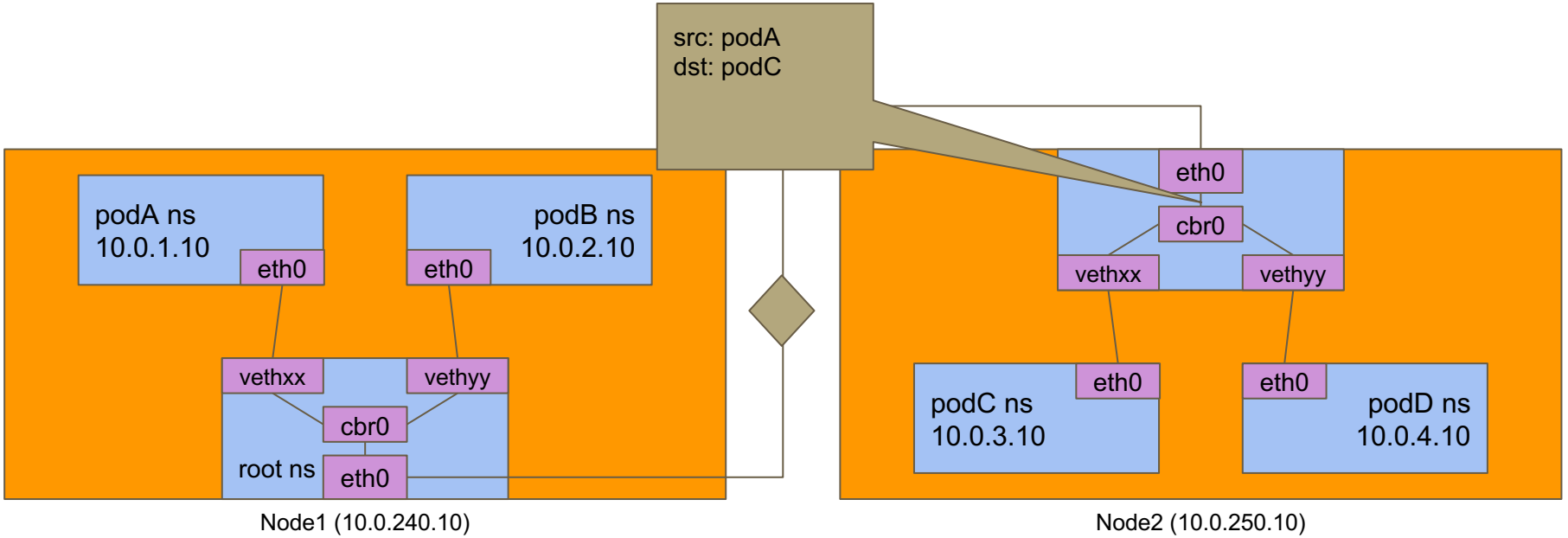
Private IP DNS name (IPv4 only)
ip-192-150-2-21.ap-south-1.compute.internal

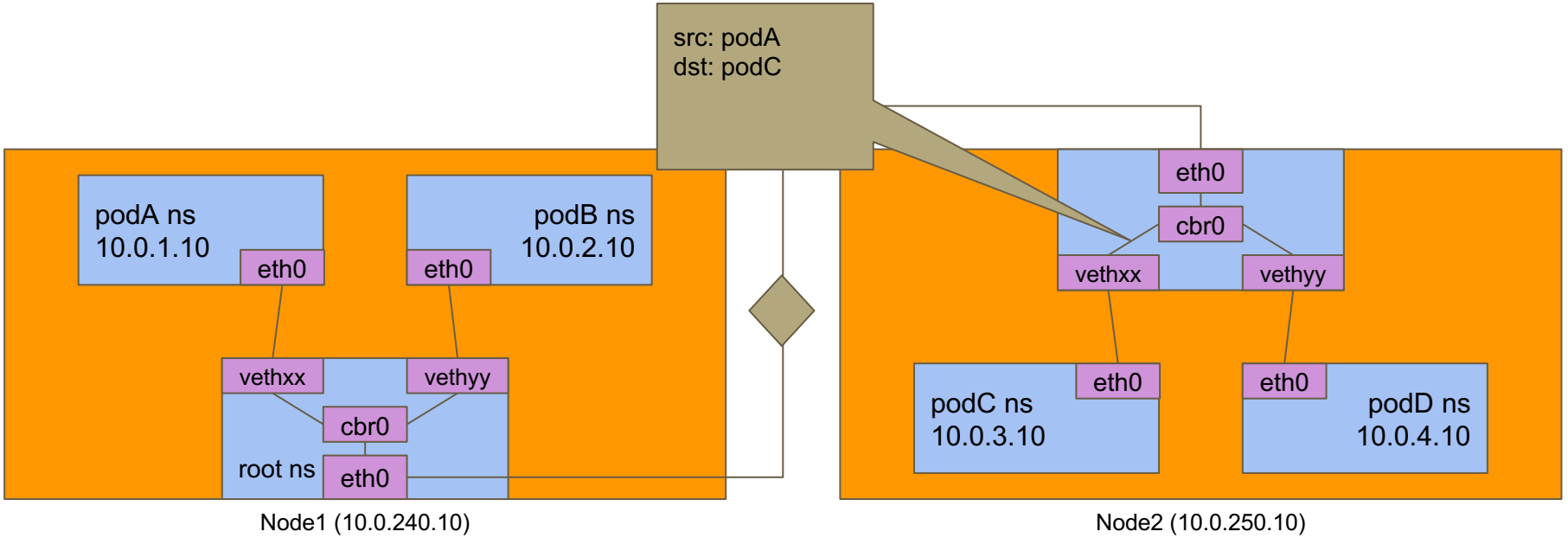
IPV6 addresses
–

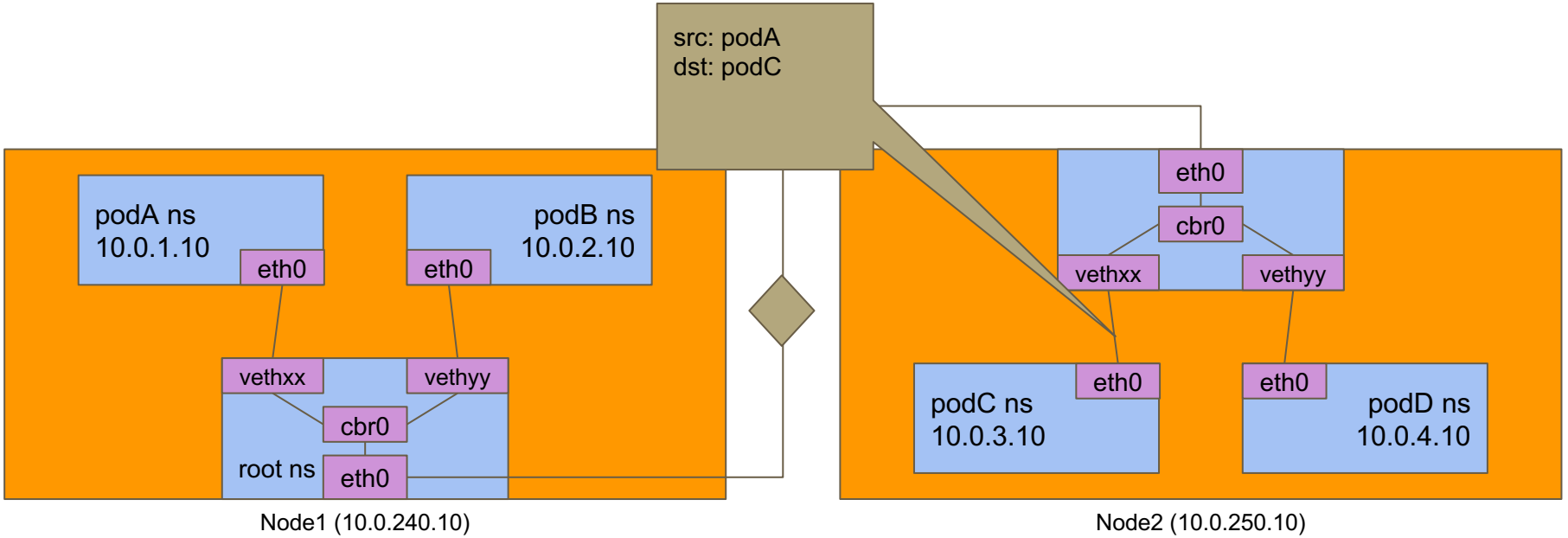
VPC ID
vpc-0fae52340ba89ac8a ([redacted]-eks-vpc)

Secondary private IPv4 addresses
192.150.2.182
192.150.2.50
192.150.2.83
192.150.2.190
192.150.2.123
192.150.2.217
192.150.2.166
192.150.2.164
192.150.2.67
192.150.2.174
192.150.2.239
192.150.2.76
192.150.2.235
192.150.2.72
192.150.2.81
192.150.2.92
192.150.2.93
192.150.2.88
192.150.2.120


src: podA
dst: podC
podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
root ns
eth0
Node1 (10.0.240.10)
eth0
cbr0
vethxx
vethyy
podC ns
10.0.3.10
eth0
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)

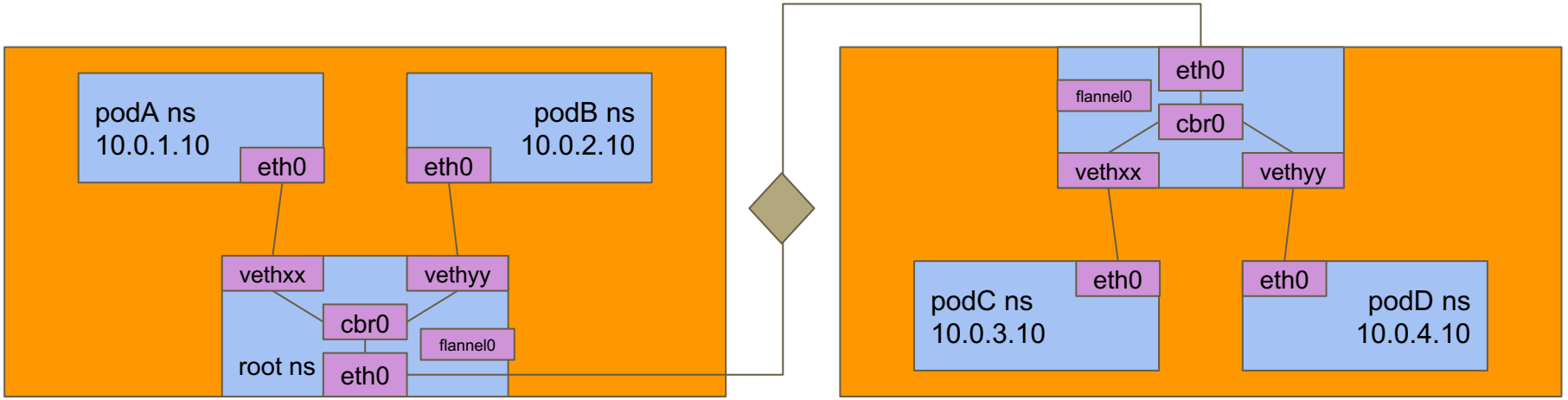

src: podA
dst: podC
podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
root ns
eth0
Node1 (10.0.240.10)
eth0
cbr0
vethxx
vethyy
podC ns
10.0.3.10
eth0
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)

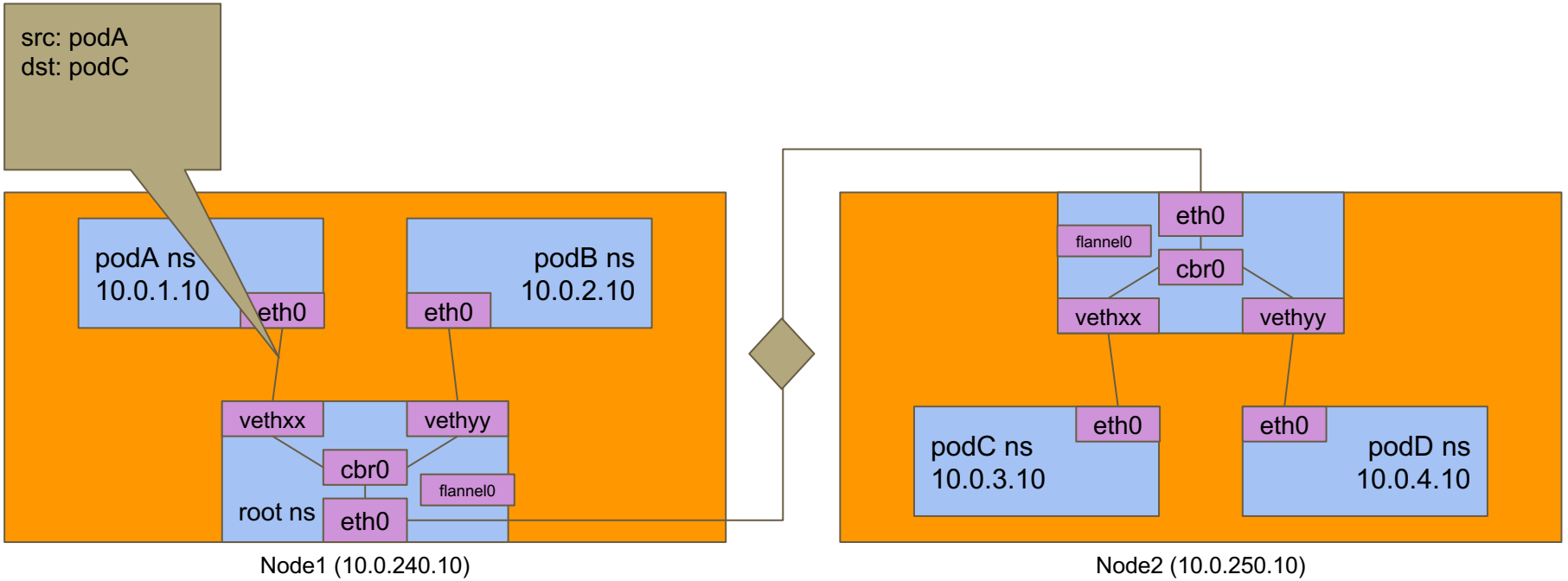

src: podA
dst: podC
podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
root ns
eth0
Node1 (10.0.240.10)
eth0
cbr0
vethxx
vethyy
podC ns
10.0.3.10
eth0
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)

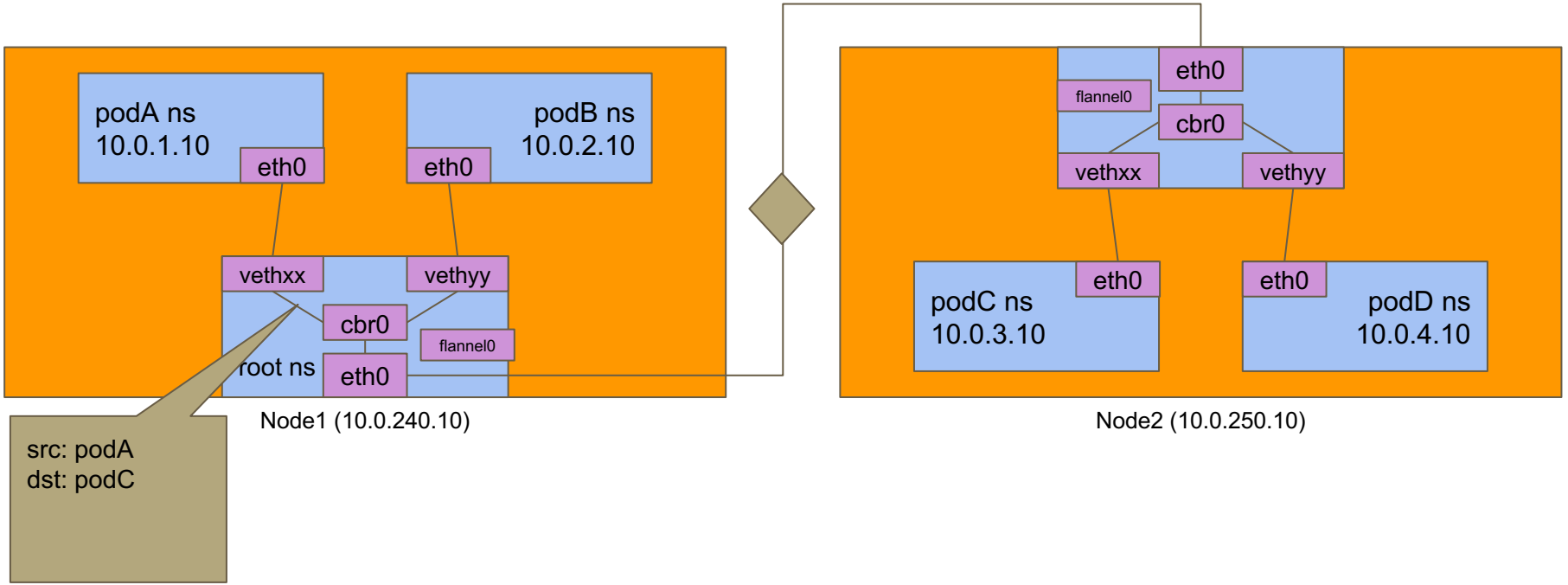

src: podA
dst: podC
podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
root ns
eth0
Node1 (10.0.240.10)
eth0
cbr0
vethxx
vethyy
podC ns
10.0.3.10
eth0
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)

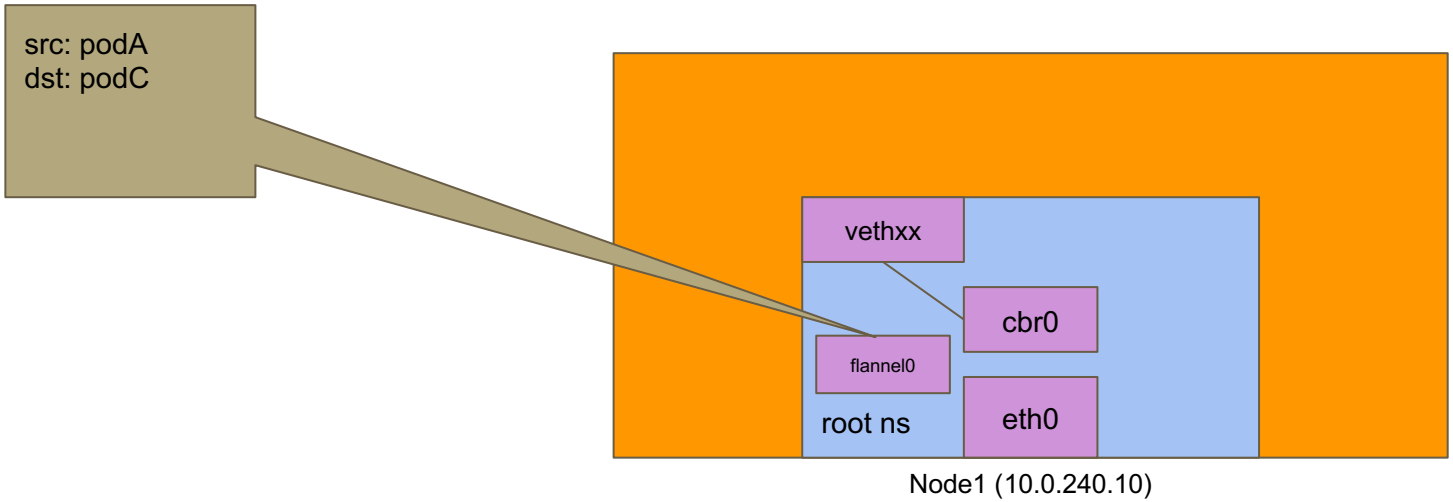
Overlay Network

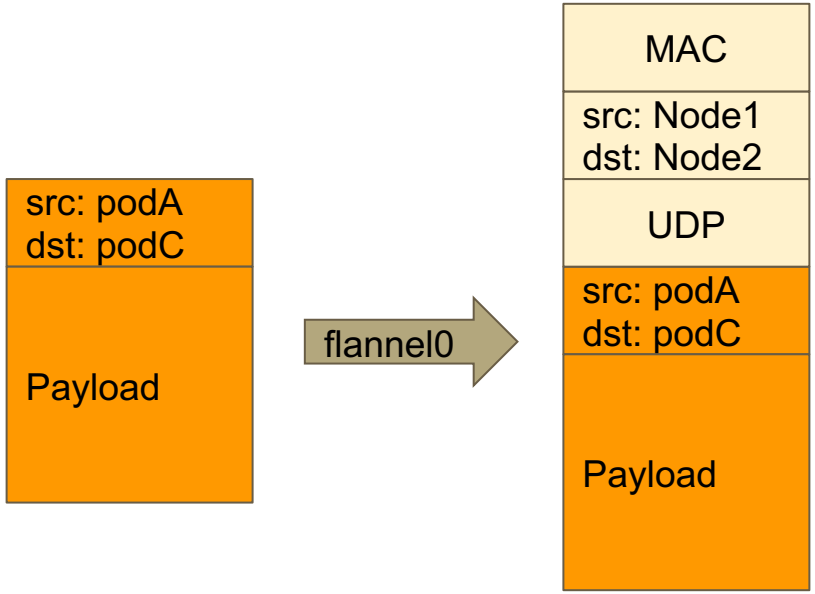
IP Ranges Gets **Exhausted**

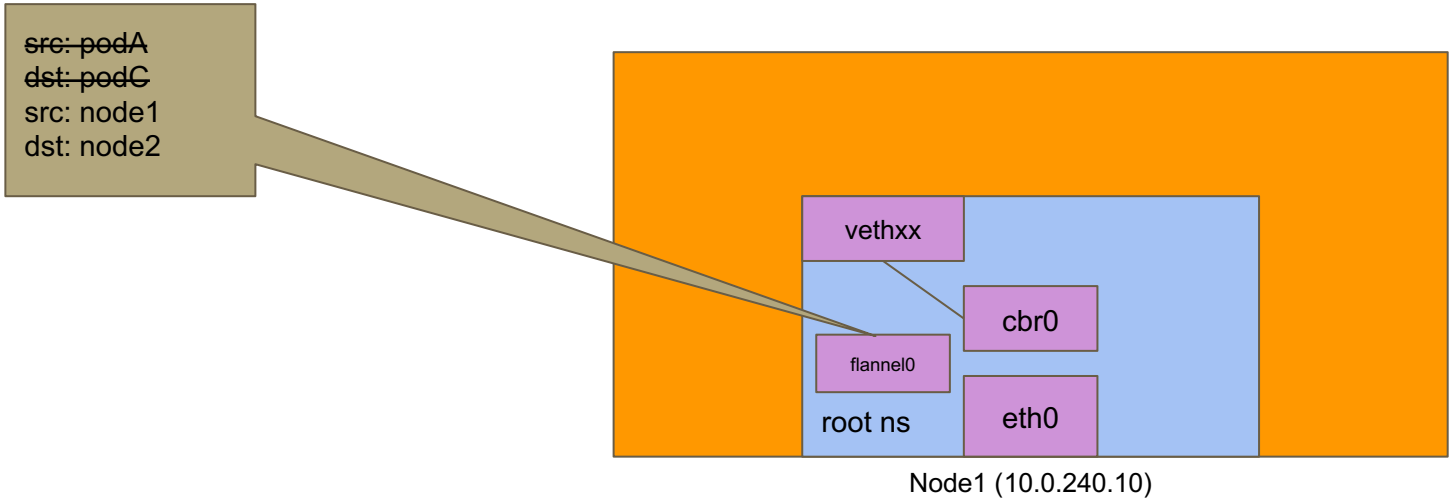

podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
flannel0
root ns
eth0
Node1 (10.0.240.10)
eth0
flannel0
cbr0
vethxx
vethyy
eth0
podC ns
10.0.3.10
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)

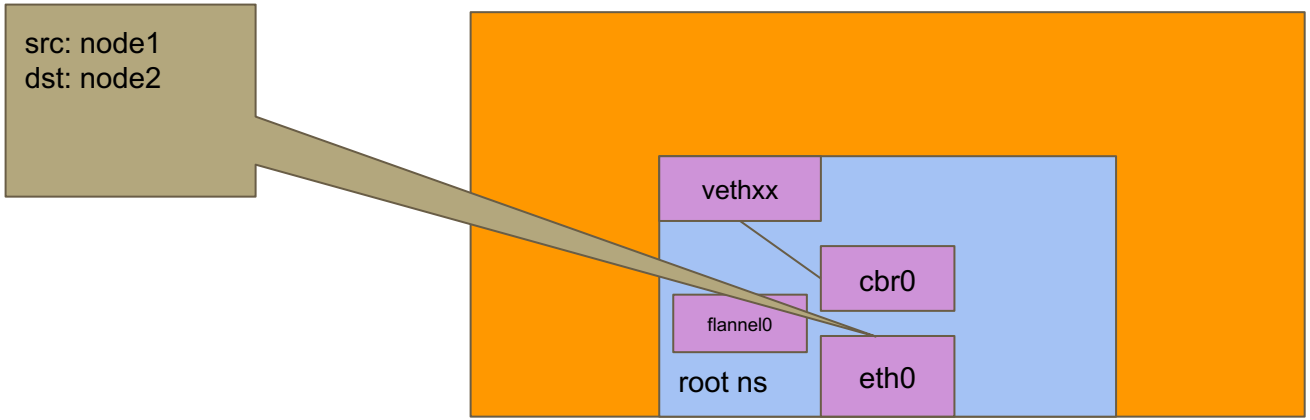

src: podA
dst: podC
podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
flannel0
root ns
eth0
Node1 (10.0.240.10)
eth0
flannel0
cbr0
vethxx
vethyy
eth0
podC ns
10.0.3.10
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)

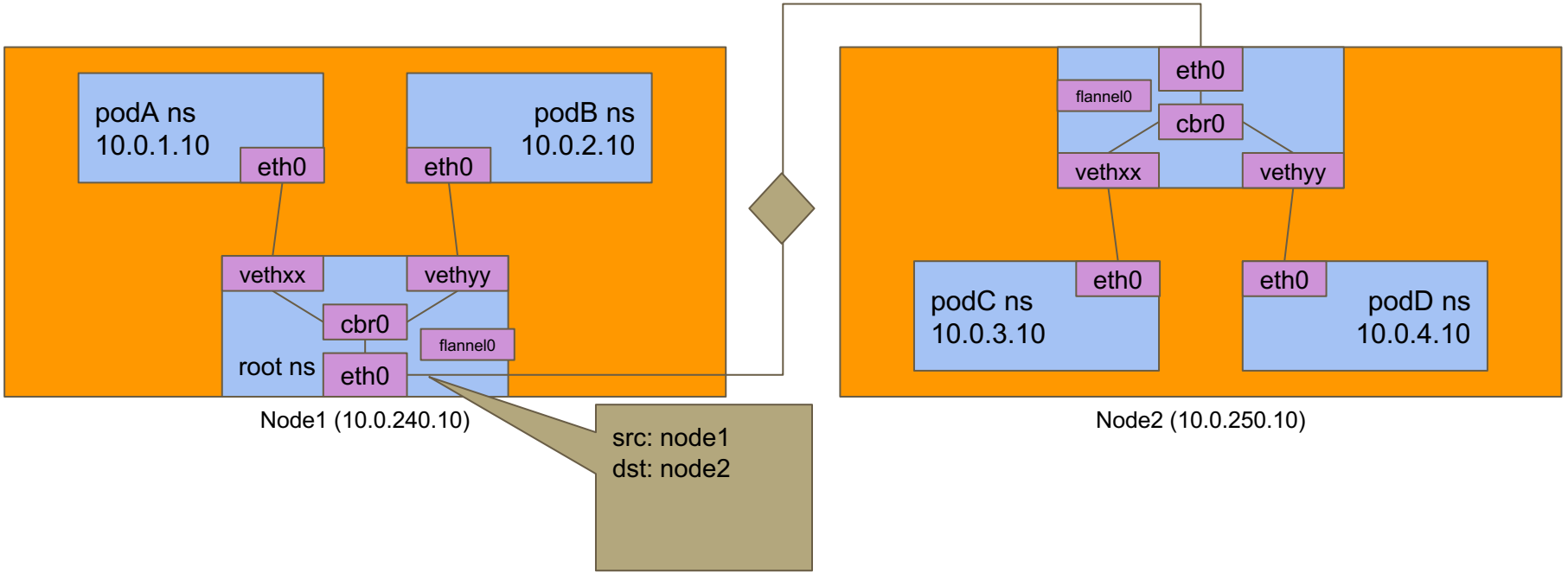

podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
flannel0
root ns
eth0
Node1 (10.0.240.10)
src: podA
dst: podC
eth0
flannel0
cbr0
vethxx
vethyy
eth0
podC ns
10.0.3.10
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)

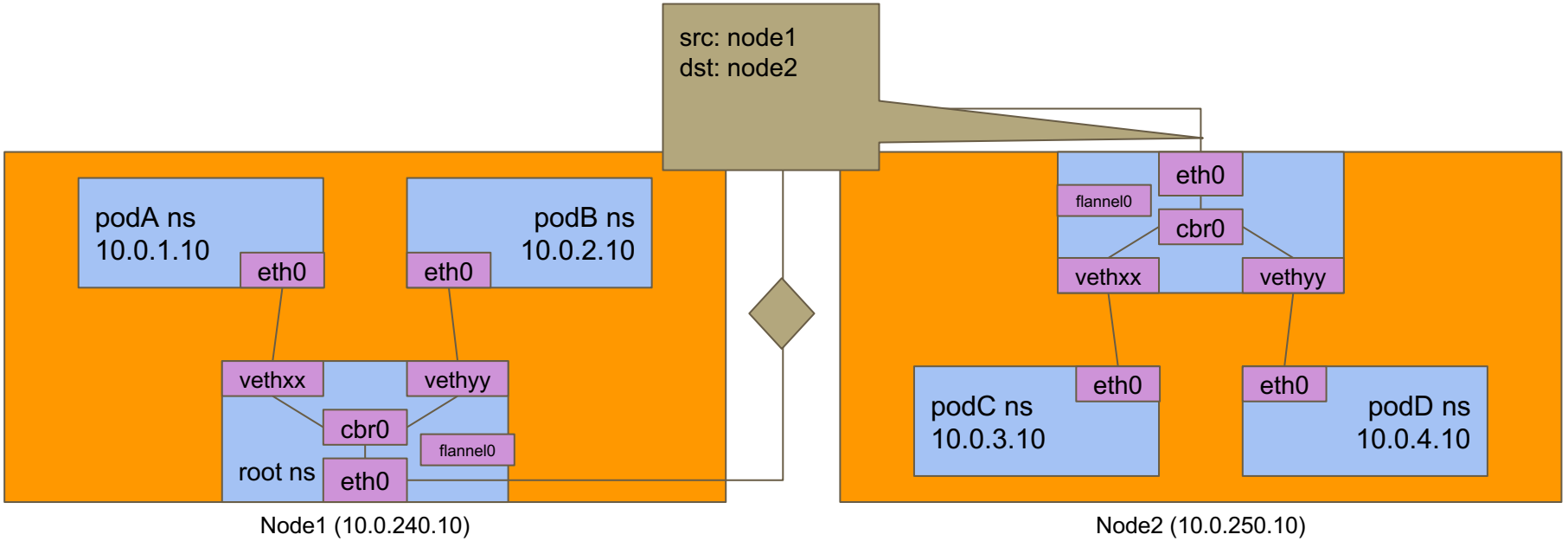

src: podA
dst: podC
vethxx
cbr0
flannel0
root ns
eth0
Node1 (10.0.240.10)

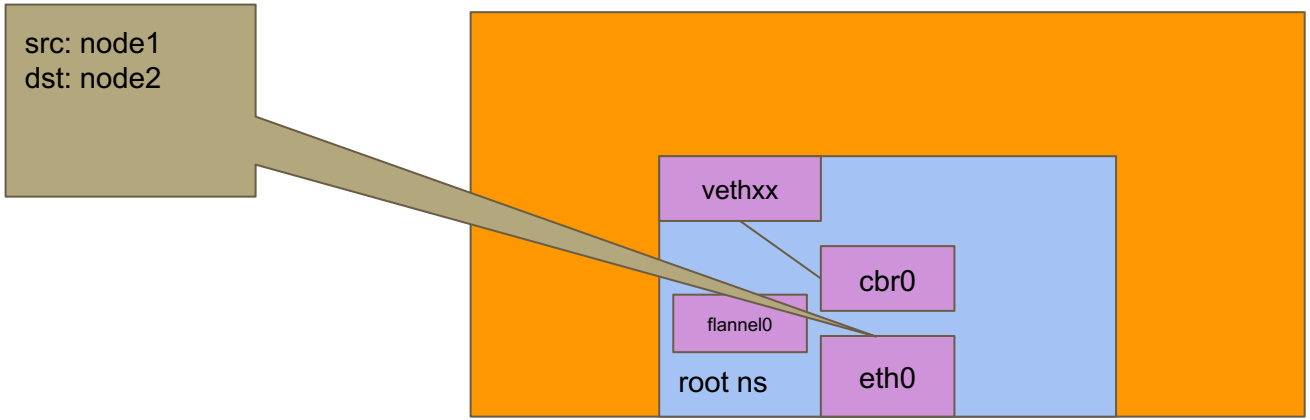

src: podA
dst: podC
Payload
flannel0
MAC
src: Node1
dst: Node2
UDP
src: podA
dst: podC
Payload

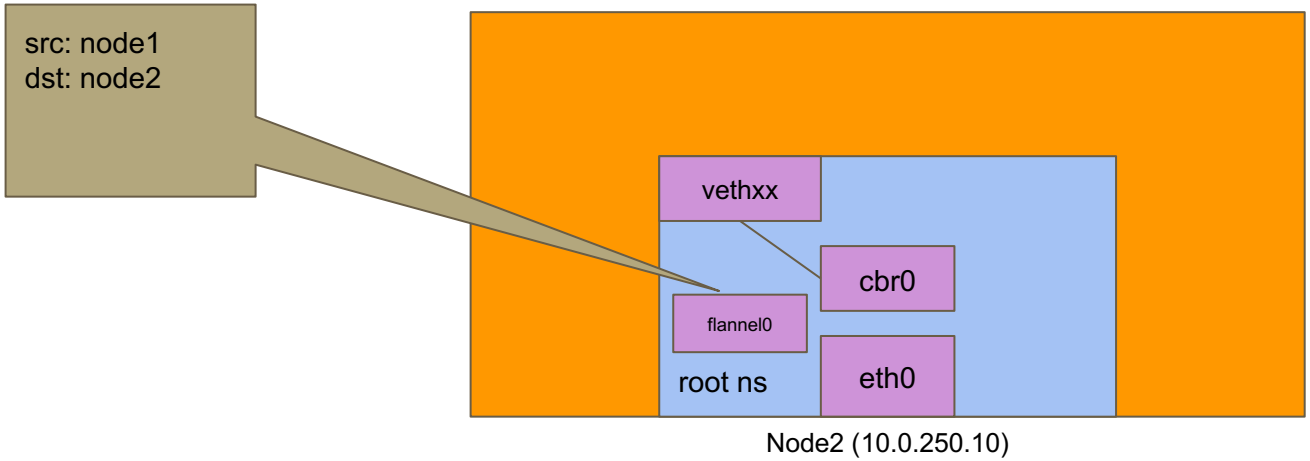

~~src: podA~~
~~dst: podC~~
src: node1
dst: node2
vethxx
cbr0
flannel0
root ns
eth0
Node1 (10.0.240.10)

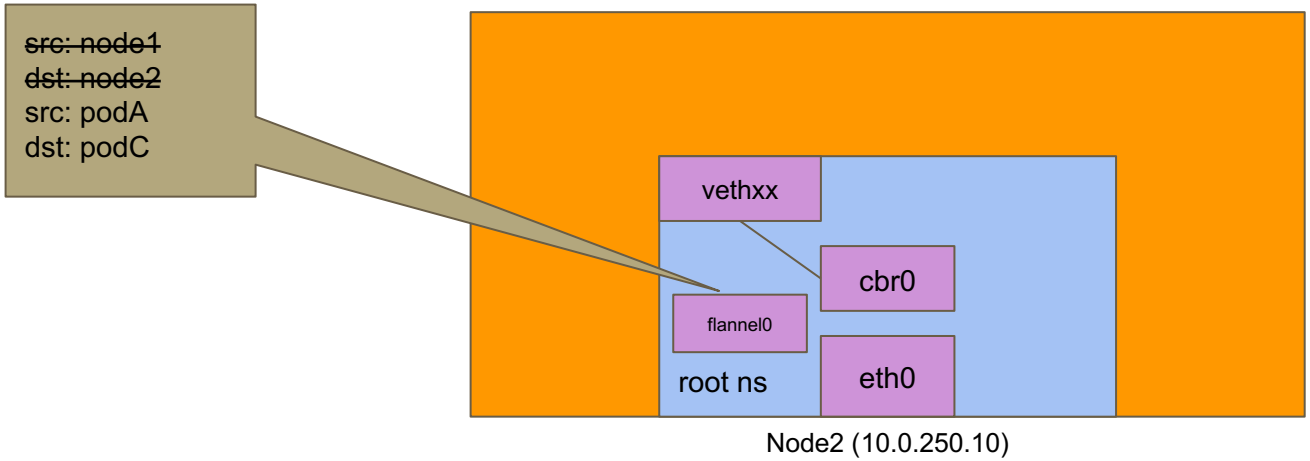

src: node1
dst: node2
vethxx
cbr0
flannel0
root ns
eth0
Node1 (10.0.240.10)

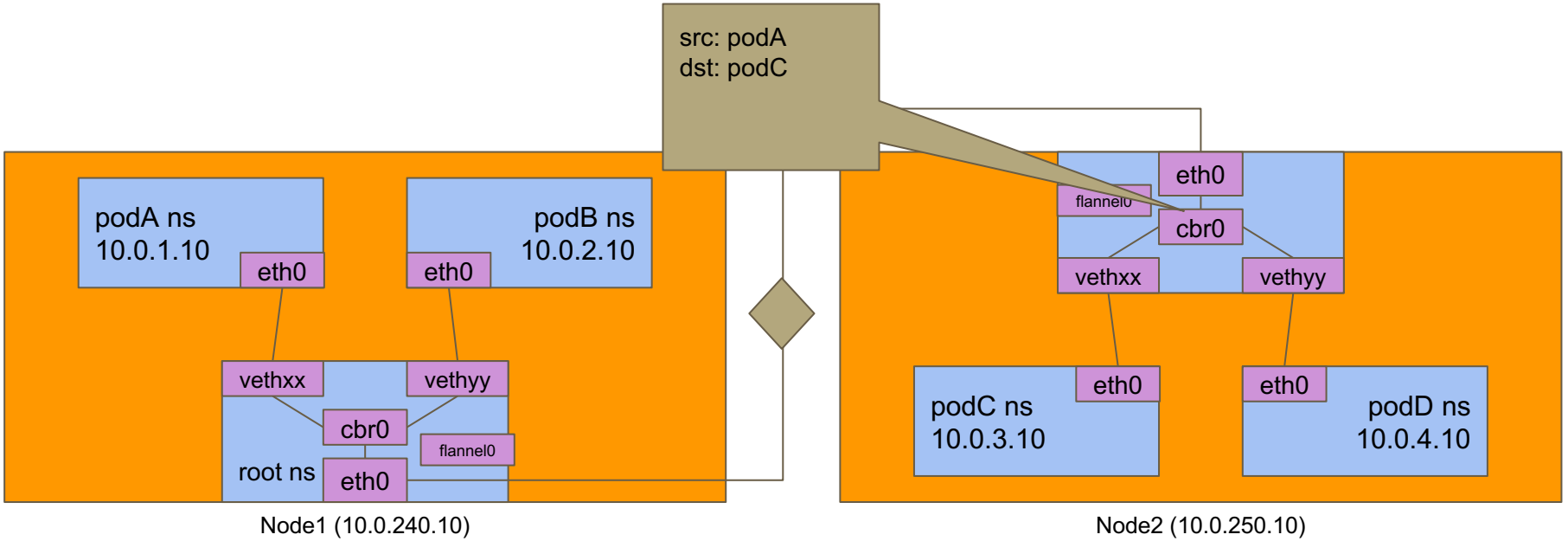

podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
flannel0
root ns
eth0
Node1 (10.0.240.10)
src: node1
dst: node2
eth0
flannel0
cbr0
vethxx
vethyy
eth0
podC ns
10.0.3.10
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)

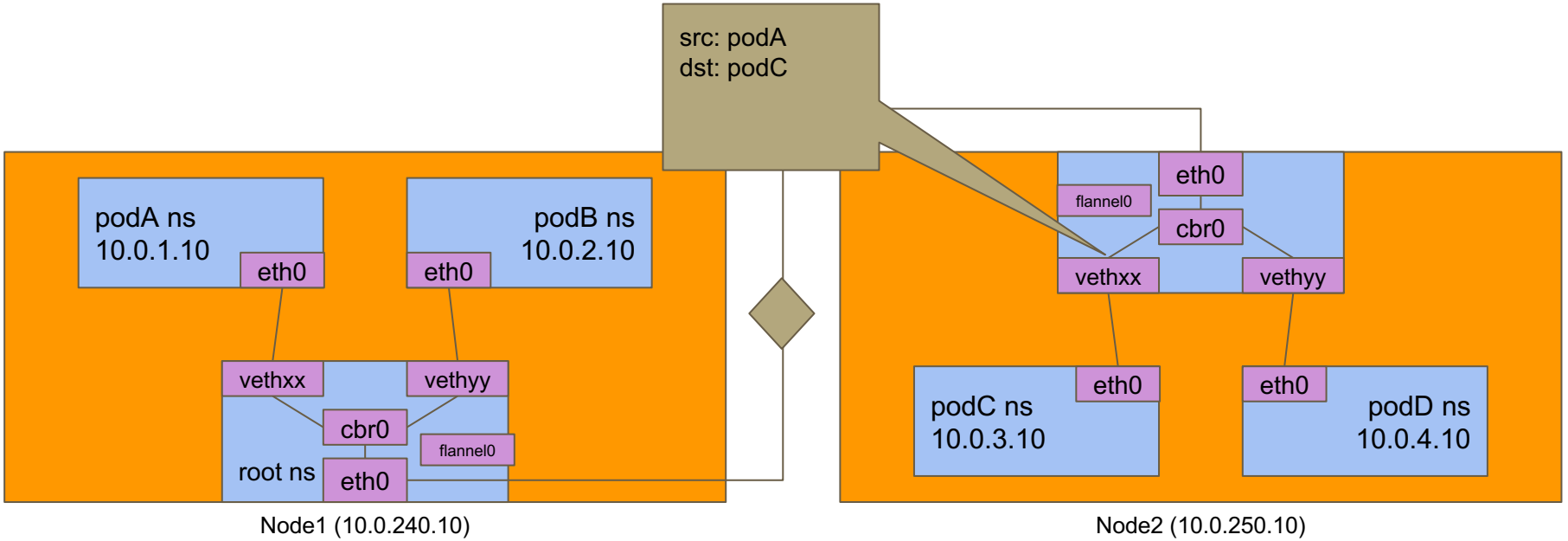

src: node1
dst: node2
podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
flannel0
root ns
eth0
Node1 (10.0.240.10)
eth0
flannel0
cbr0
vethxx
vethyy
podC ns
10.0.3.10
eth0
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)

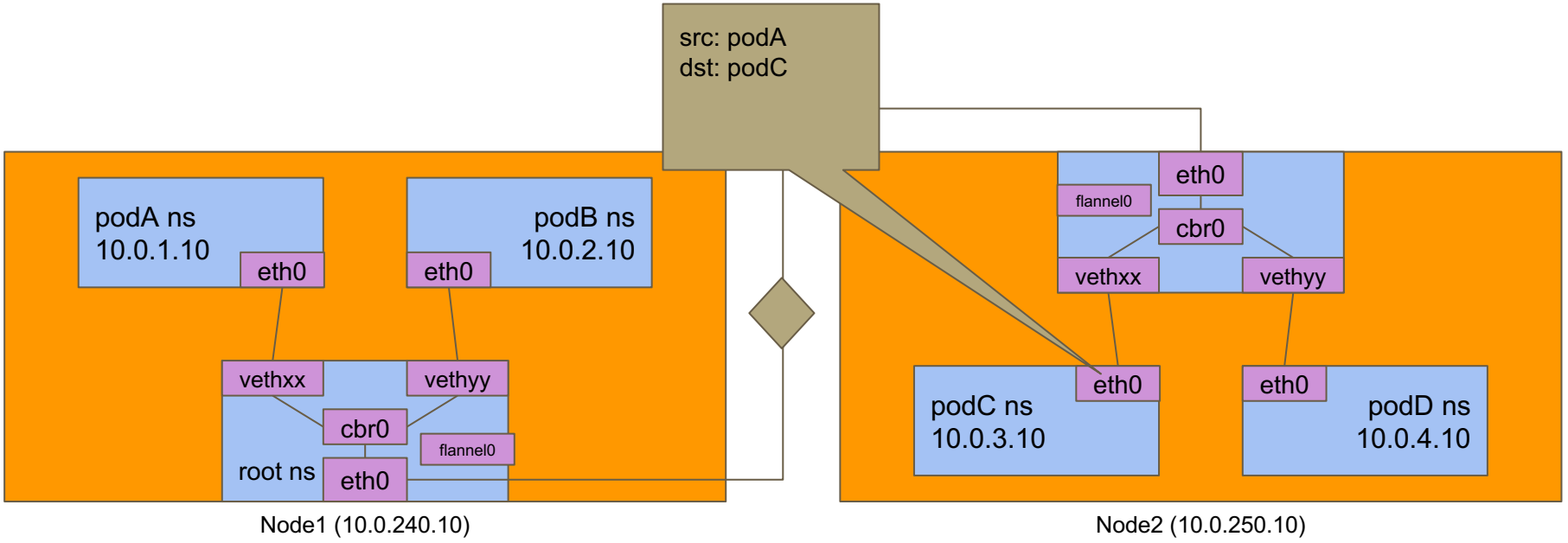

src: node1
dst: node2
vethxx
cbr0
flannel0
root ns
eth0
Node2 (10.0.250.10)


src: node1
dst: node2
vethxx
cbr0
flannel0
root ns
eth0
Node2 (10.0.250.10)


~~src: node1~~
~~dst: node2~~
src: podA
dst: podC
vethxx
cbr0
flannel0
root ns
eth0
Node2 (10.0.250.10)


src: podA
dst: podC
podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
flannel0
root ns
eth0
Node1 (10.0.240.10)
eth0
flannel0
cbr0
vethxx
vethyy
podC ns
10.0.3.10
eth0
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)


src: podA
dst: podC
podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
flannel0
root ns
eth0
Node1 (10.0.240.10)
eth0
flannel0
cbr0
vethxx
vethyy
podC ns
10.0.3.10
eth0
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)


src: podA
dst: podC
podA ns
10.0.1.10
eth0
podB ns
10.0.2.10
eth0
vethxx
vethyy
cbr0
flannel0
root ns
eth0
Node1 (10.0.240.10)
eth0
flannel0
cbr0
vethxx
vethyy
eth0
podC ns
10.0.3.10
eth0
podD ns
10.0.4.10
Node2 (10.0.250.10)

What **If**

The Pod **IP**

Address **Changes**?

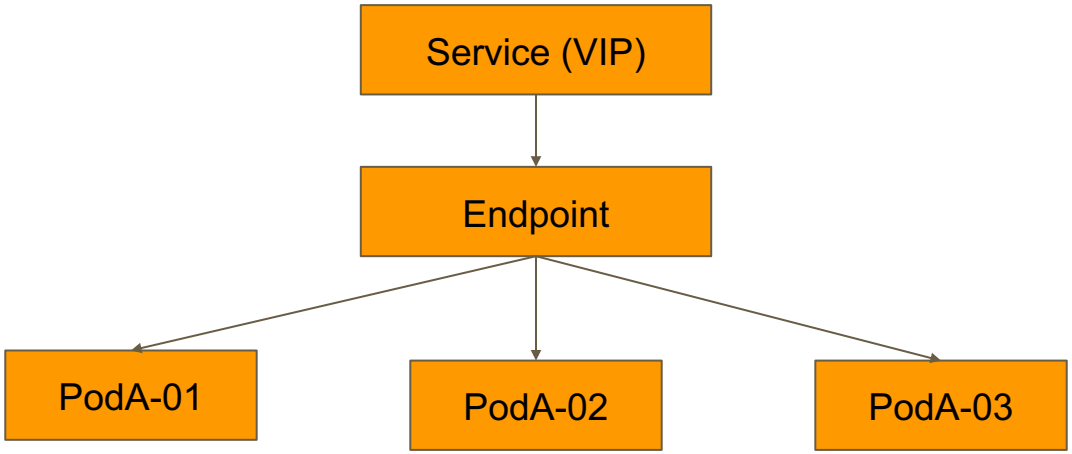
Service

Abstraction

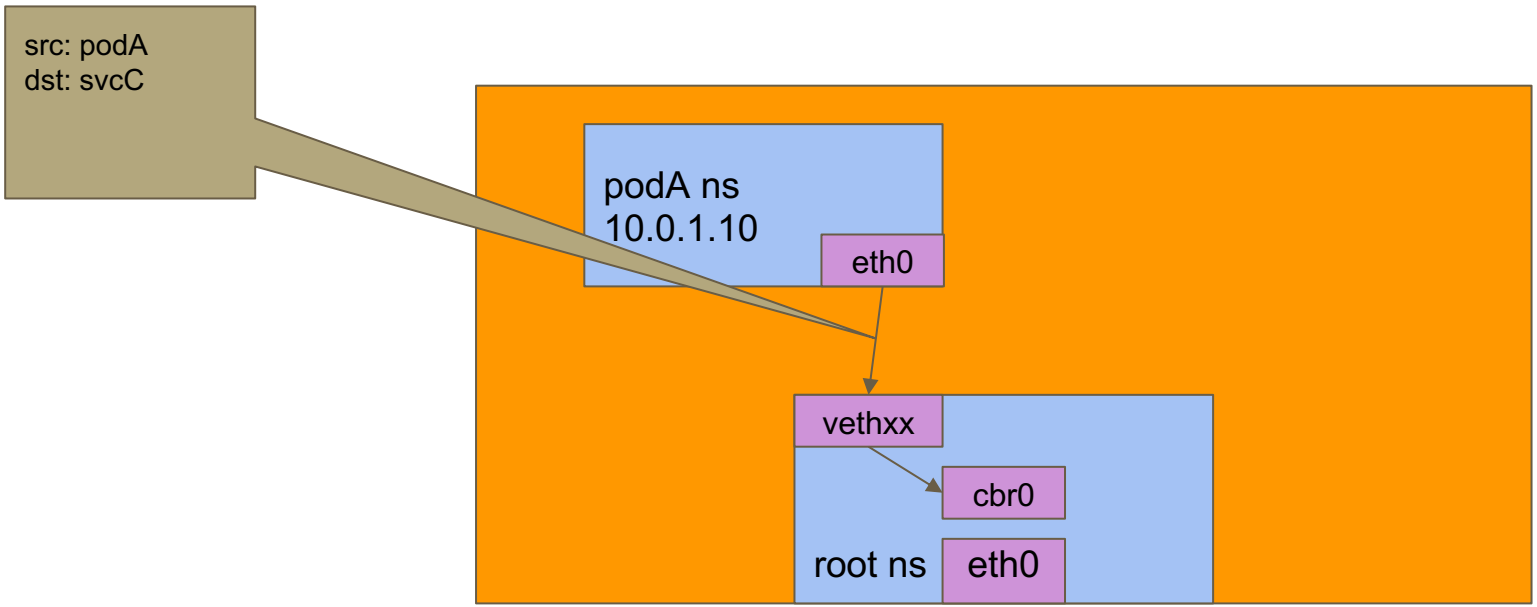
For A

Group Of Pods

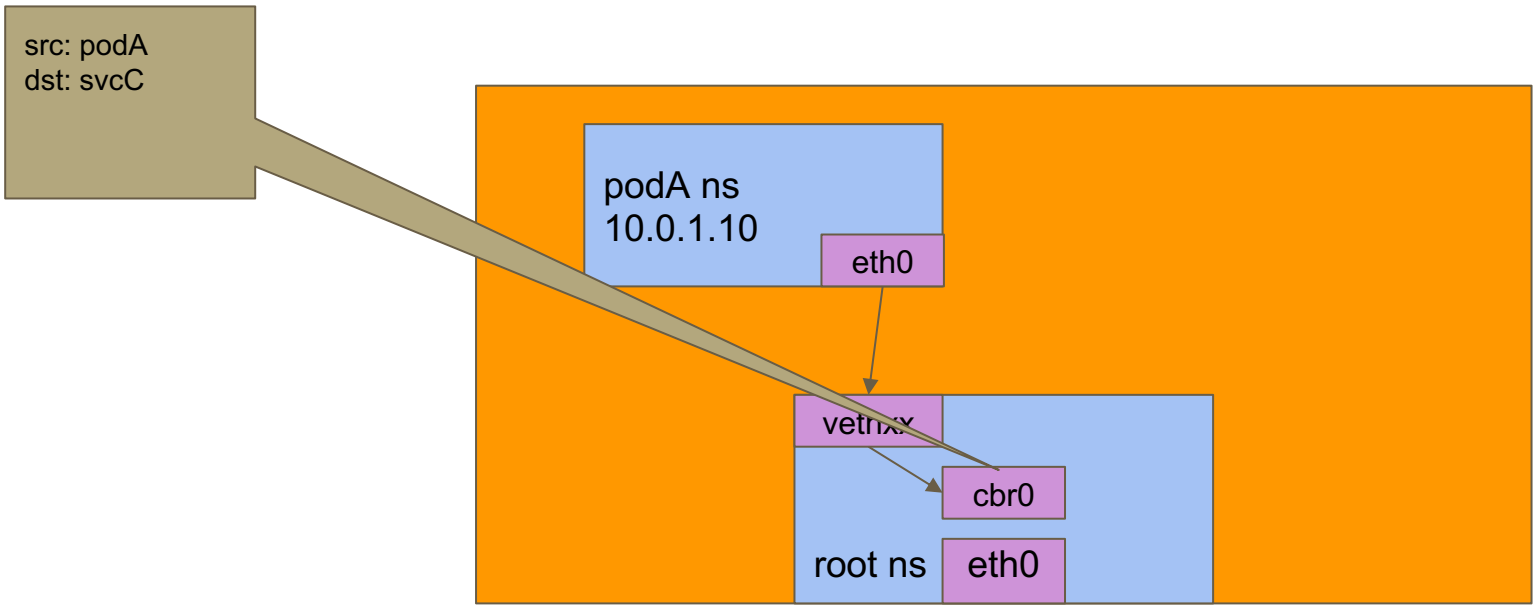
```
apiVersion: v1
kind: Service

metadata:
  name: hello-kubernetes

spec:
  ports:
  - port: 80
    protocol: TCP
    targetPort: 8080
  selector:
    app: hello-kubernetes
  type: ClusterIP
```


Service (VIP)
Endpoint
PodA-01
PodA-02
PodA-03

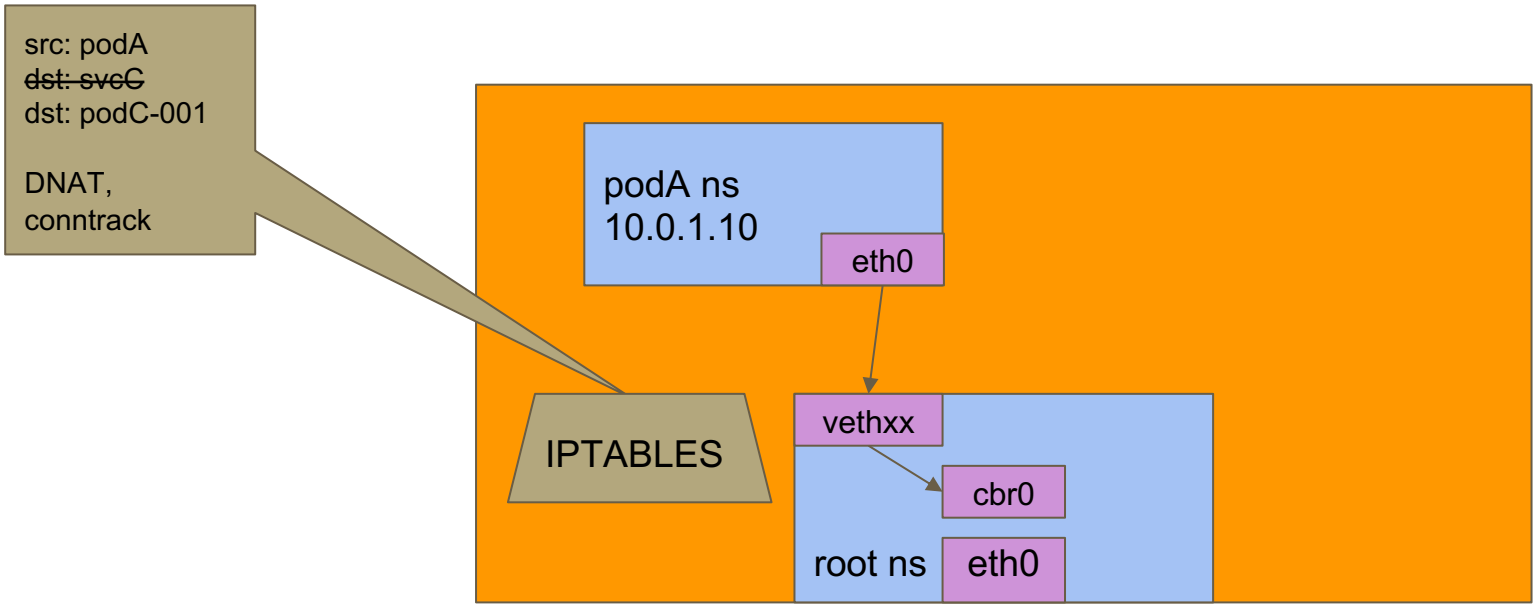
How Does A Pod Talk To Another Pod Using Service IP?



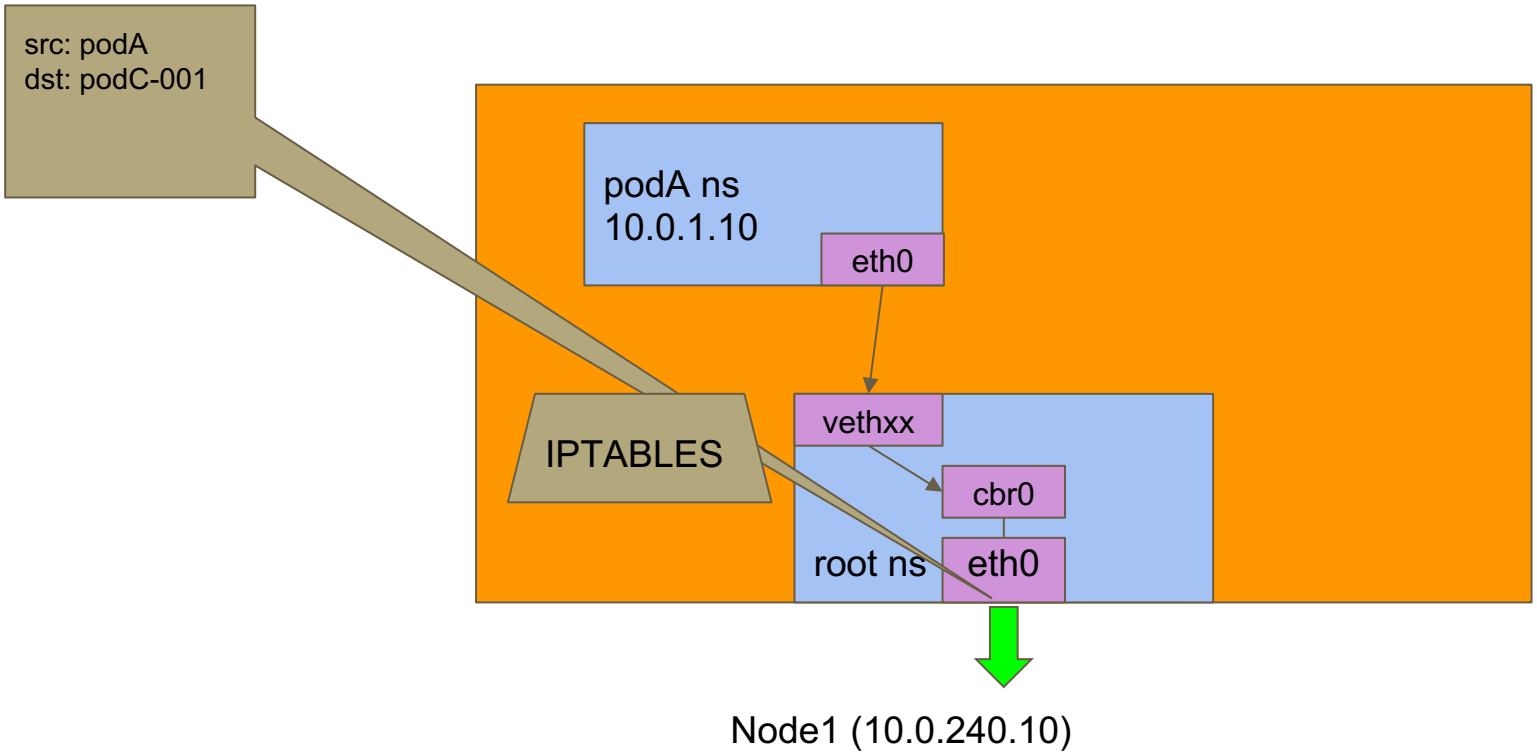
Node1 (10.0.240.10)

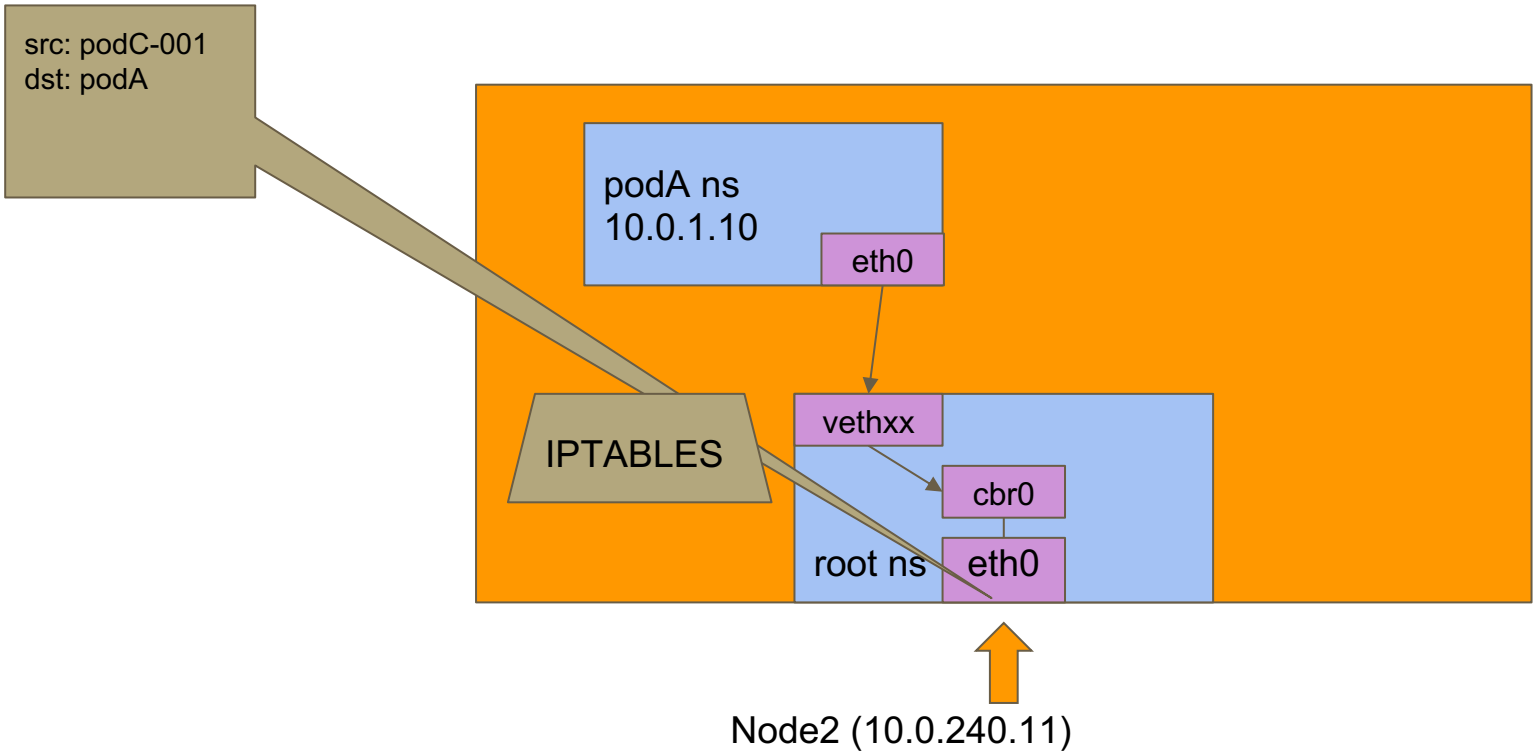


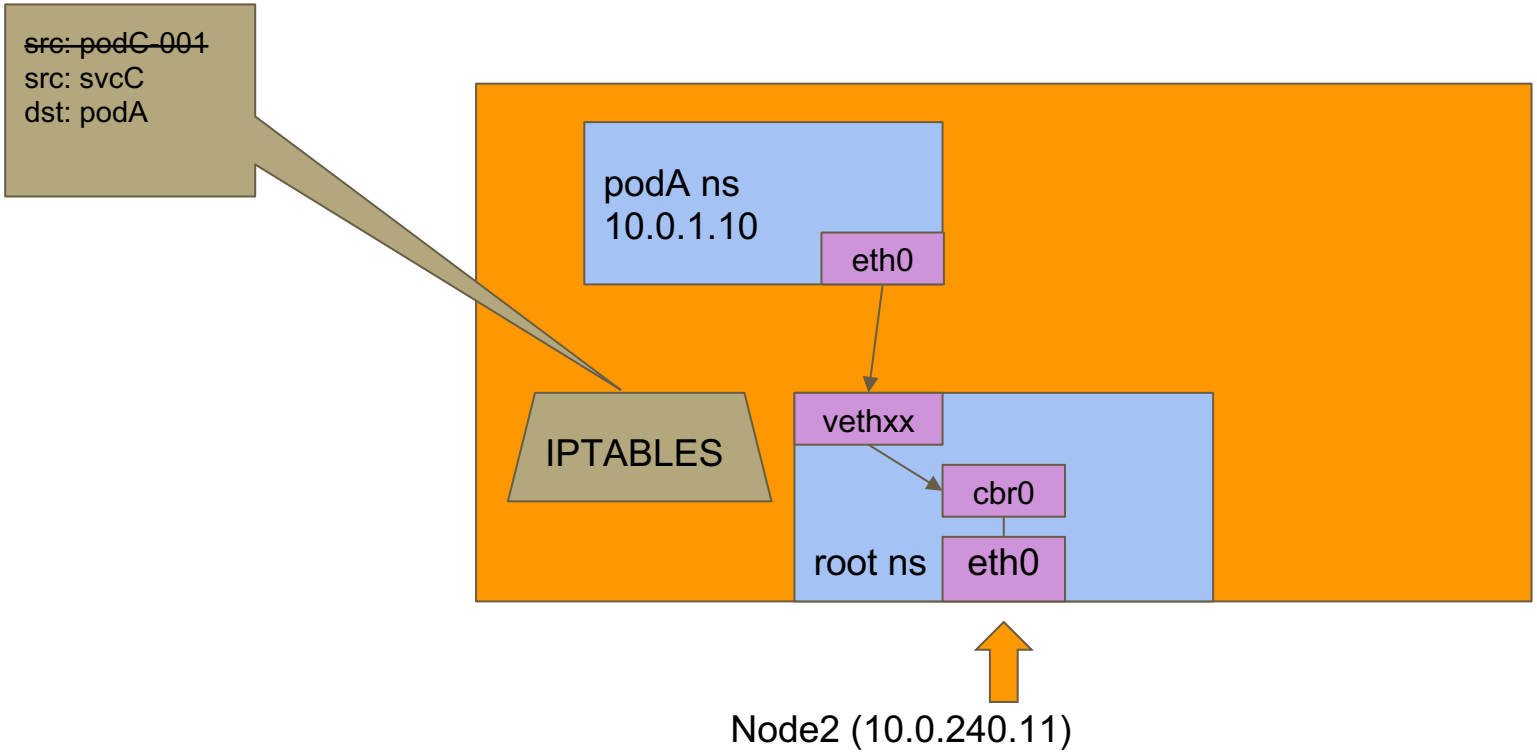
Node1 (10.0.240.10)

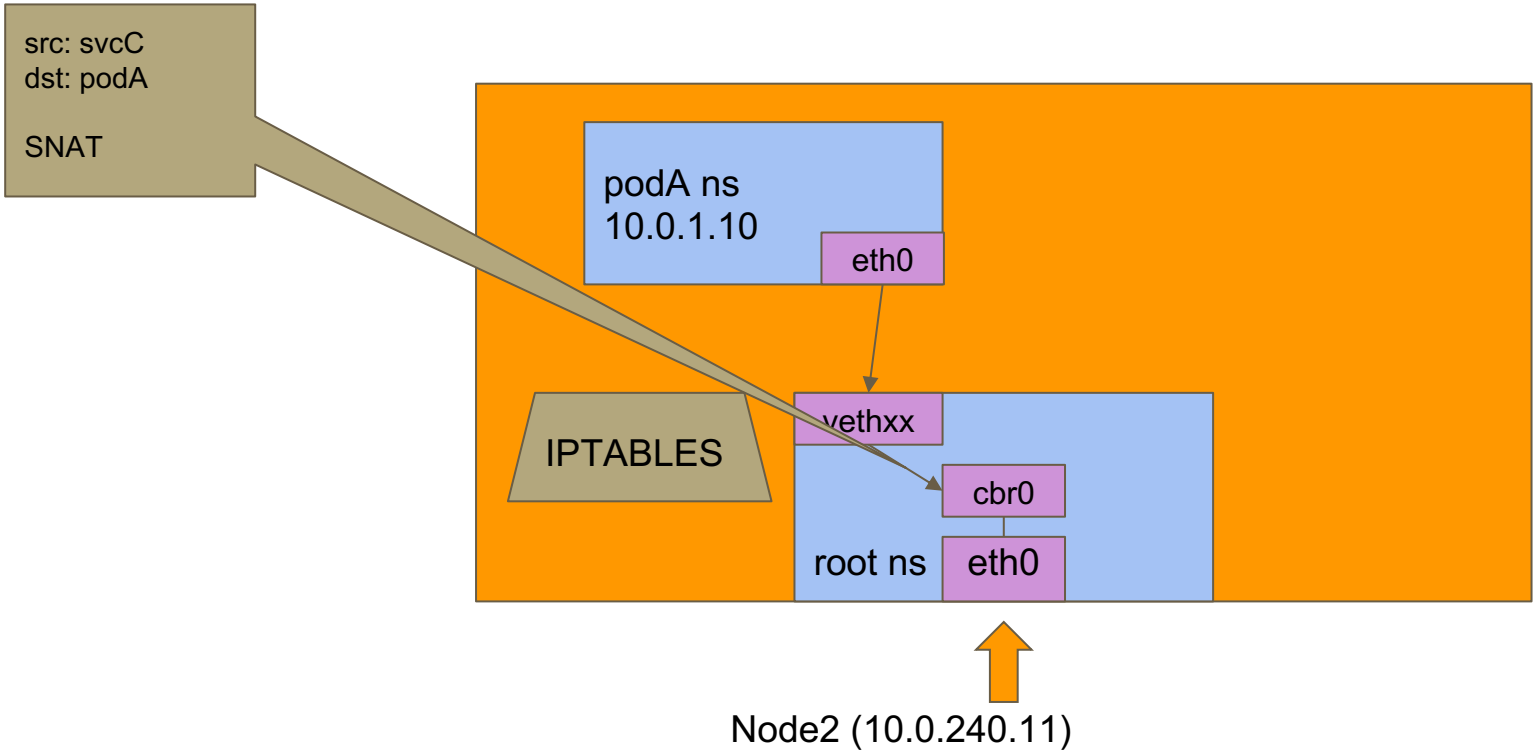


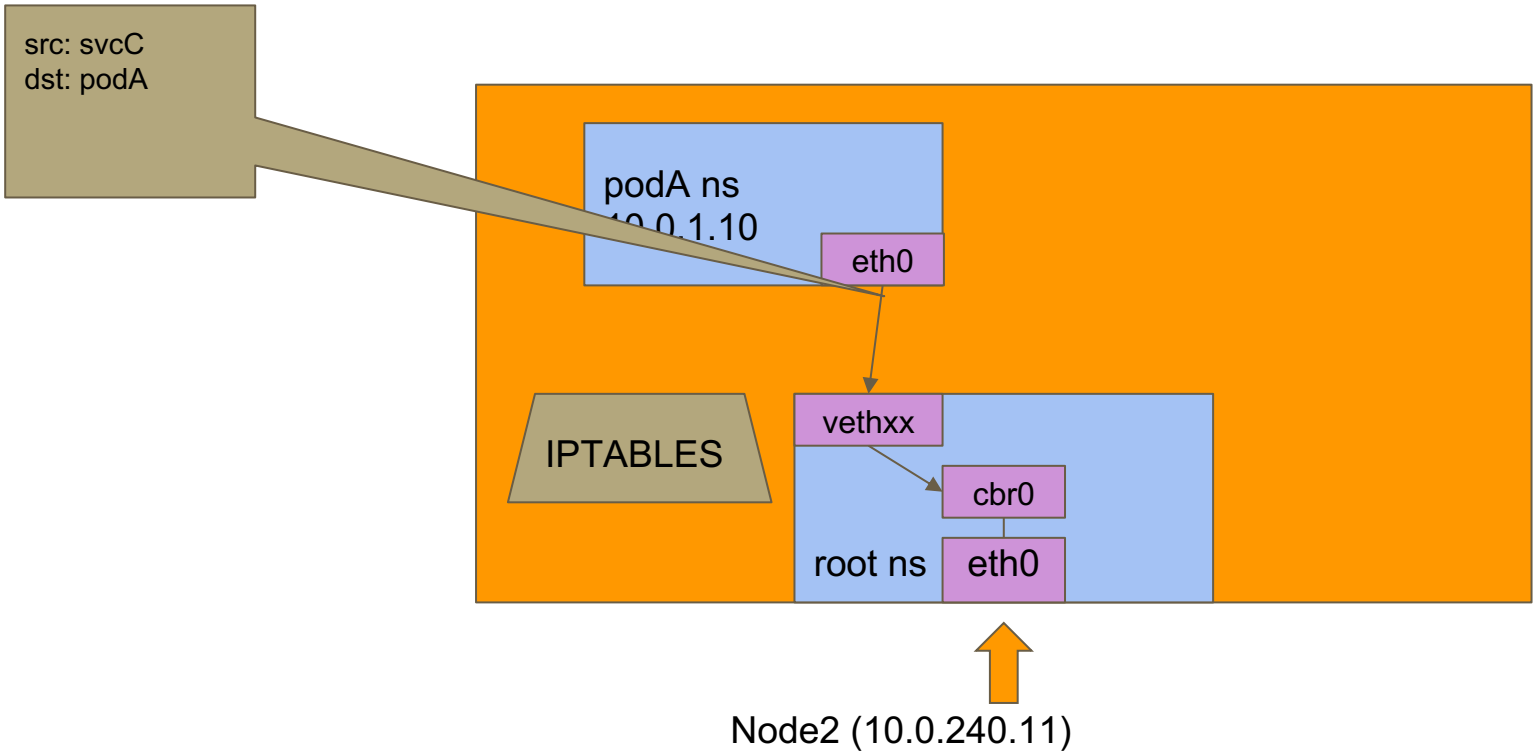
Node1 (10.0.240.10)


src: podA
dst: podC-001
podA ns
10.0.1.10
eth0
IPTABLES
vethxx
cbr0
root ns
eth0
Node1 (10.0.240.10)


src: podC-001
dst: podA
podA ns
10.0.1.10
eth0
IPTABLES
vethxx
cbr0
root ns
eth0
Node2 (10.0.240.11)


~~src: podC-001~~
src: svcC
dst: podA
podA ns
10.0.1.10
eth0
IPTABLES
vethxx
cbr0
root ns
eth0
Node2 (10.0.240.11)


src: svcC
dst: podA
SNAT
podA ns
10.0.1.10
eth0
IPTABLES
vethxx
cbr0
root ns
eth0
Node2 (10.0.240.11)


src: svcC
dst: podA
podA ns
eth0
IPTABLES
vethxx
cbr0
root ns
eth0
Node2 (10.0.240.11)

```
if dest.ip == service-a.ip {

        pick one of the pods for service-a &&
        rewrite the packet

}
```

QnA

THANK YOU!

Resources

https://learncloudnative.com/blog/2023-05-31-kubeproxy-iptables

https://www.redhat.com/sysadmin/kubernetes-pods-communicate-nodes

https://www.youtube.com/watch?v=0Omvgd7Hg1I

https://www.youtube.com/watch?v=InZVNuKY5GY&t=867s