Copyright And Confidentiality

羅偉

牧云/罗伟
Network | IaaS | PaaS | ServiceMesh
交流 学习 沉淀 成长 分享
olaf.luo@foxmail.com
https://www.yuque.com/wei.luo
https://youdianzhishi.com

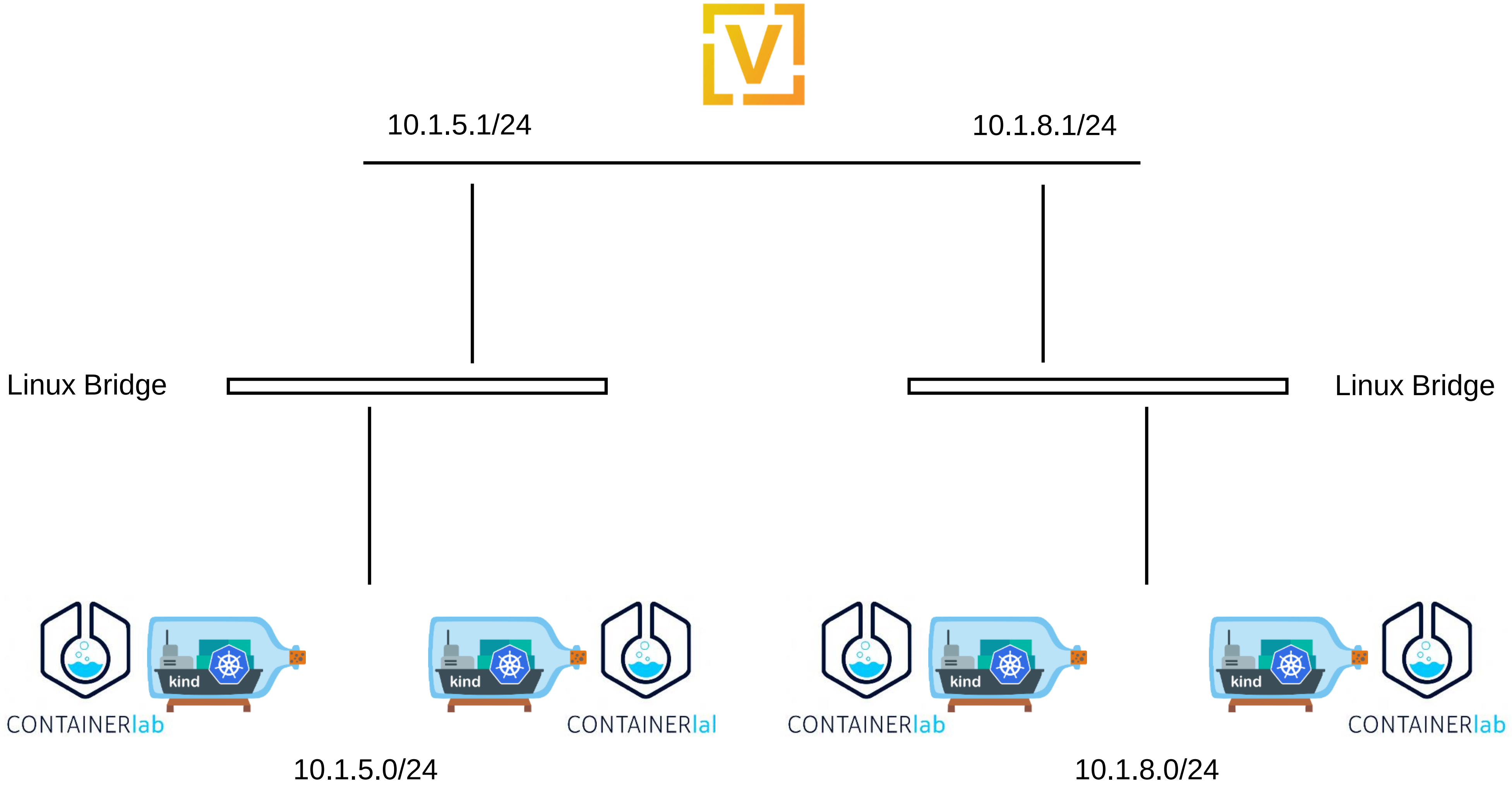
Rowan Luo

交流 学习 沉淀 成长 分享

Kubernetes CNI - Calico IPIP CrossSubnet Mode

Revision Date: 2023/02/23
Version：v1.2
DocID：CN00002023XLW

Calico IPIP CrossSubnet Mode - Prepare Environment



10.1.5.1/24

10.1.8.1/24

Linux Bridge

Linux Bridge

10.1.5.0/24

10.1.8.0/24

Calico IPIP CrossSubnet Mode - Prepare Environment

```
# 1.prep noCNI env
cat <<EOF | kind create cluster --name=clab-calico-ipip-crosssubnet --image=kindest/node:v1.23.4 --config=-
kind: Cluster
apiVersion: kind.x-k8s.io/v1alpha4
networking:
  disableDefaultCNI: true
  podSubnet: "10.244.0.0/16"
nodes:
- role: control-plane
  kubeadmConfigPatches:
  - |
    kind: InitConfiguration
    nodeRegistration:
      kubeletExtraArgs:
        node-ip: 10.1.5.10
        node-labels: "rack=rack0"
- role: worker
  kubeadmConfigPatches:
  - |
    kind: JoinConfiguration
    nodeRegistration:
      kubeletExtraArgs:
        node-ip: 10.1.5.11
        node-labels: "rack=rack0"
- role: worker
  kubeadmConfigPatches:
  - |
    kind: JoinConfiguration
    nodeRegistration:
      kubeletExtraArgs:
        node-ip: 10.1.8.10
        node-labels: "rack=rack1"
- role: worker
  kubeadmConfigPatches:
  - |
    kind: JoinConfiguration
    nodeRegistration:
      kubeletExtraArgs:
        node-ip: 10.1.8.11
        node-labels: "rack=rack1"
containerdConfigPatches:
- |-
  [plugins."io.containerd.grpc.v1.cri".registry.mirrors."192.168.2.100:5000"]
    endpoint = ["http://192.168.2.100:5000"]
EOF
```

Calico IPIP CrossSubnet Mode - Prepare Environment

```
name: calico-ipip-crosssubnet
topology:
  nodes:
    gw0:
      kind: linux
      image: 192.168.2.100:5000/vyos/vyos:1.2.8
      cmd: /sbin/init
      binds:
        - /lib/modules:/lib/modules
        - ./startup-conf/gw0-boot.cfg:/opt/vyatta/etc/config/config.boot
    br-pool0:
      kind: bridge
    br-pool1:
      kind: bridge
    server1:
      kind: linux
      image: 192.168.2.100:5000/nettool
      network-mode: container:control-plane
      exec:
      - ip addr add 10.1.5.10/24 dev net0
      - ip route replace default via 10.1.5.1
    server2:
      kind: linux
      image: 192.168.2.100:5000/nettool
      network-mode: container:worker
      exec:
      - ip addr add 10.1.5.11/24 dev net0
      - ip route replace default via 10.1.5.1
    server3:
      kind: linux
      image: 192.168.2.100:5000/nettool
      network-mode: container:worker2
      exec:
      - ip addr add 10.1.8.10/24 dev net0
      - ip route replace default via 10.1.8.1
    server4:
      kind: linux
      image: 192.168.2.100:5000/nettool
      network-mode: container:worker3
      exec:
      - ip addr add 10.1.8.11/24 dev net0
      - ip route replace default via 10.1.8.1
  links:
    - endpoints: ["br-pool0:br-pool0-net0", "server1:net0"]
    - endpoints: ["br-pool0:br-pool0-net1", "server2:net0"]
    - endpoints: ["br-pool1:br-pool1-net0", "server3:net0"]
    - endpoints: ["br-pool1:br-pool1-net1", "server4:net0"]
    - endpoints: ["gw0:eth1", "br-pool0:br-pool0-net2"]
    - endpoints: ["gw0:eth2", "br-pool1:br-pool1-net2"]
```

Calico IPIP CrossSubnet Mode - IPIP Basics

In the most general tunneling case we have:

source ---> encapsulator --------> decapsulator ---> destination

with the source, encapsulator, decapsulator, and destination being separate nodes. The encapsulator node is considered the "entry point" of the tunnel, and the decapsulator node is considered the "exit point" of the tunnel. There in general may be multiple source-destination pairs using the same tunnel between the encapsulator and decapsulator.

An IP tunnel is an Internet Protocol (IP) network communications channel between two networks. It is used to transport another network protocol by encapsulation of its packets.

IP tunnels are often used for connecting two disjoint IP networks that don't have a native routing path to each other, via an underlying routable protocol across an intermediate transport network. In conjunction with the IPsec protocol they may be used to create a virtual private network between two or more private networks across a public network such as the Internet. Another prominent use is to connect islands of IPv6 installations across the IPv4 Internet.

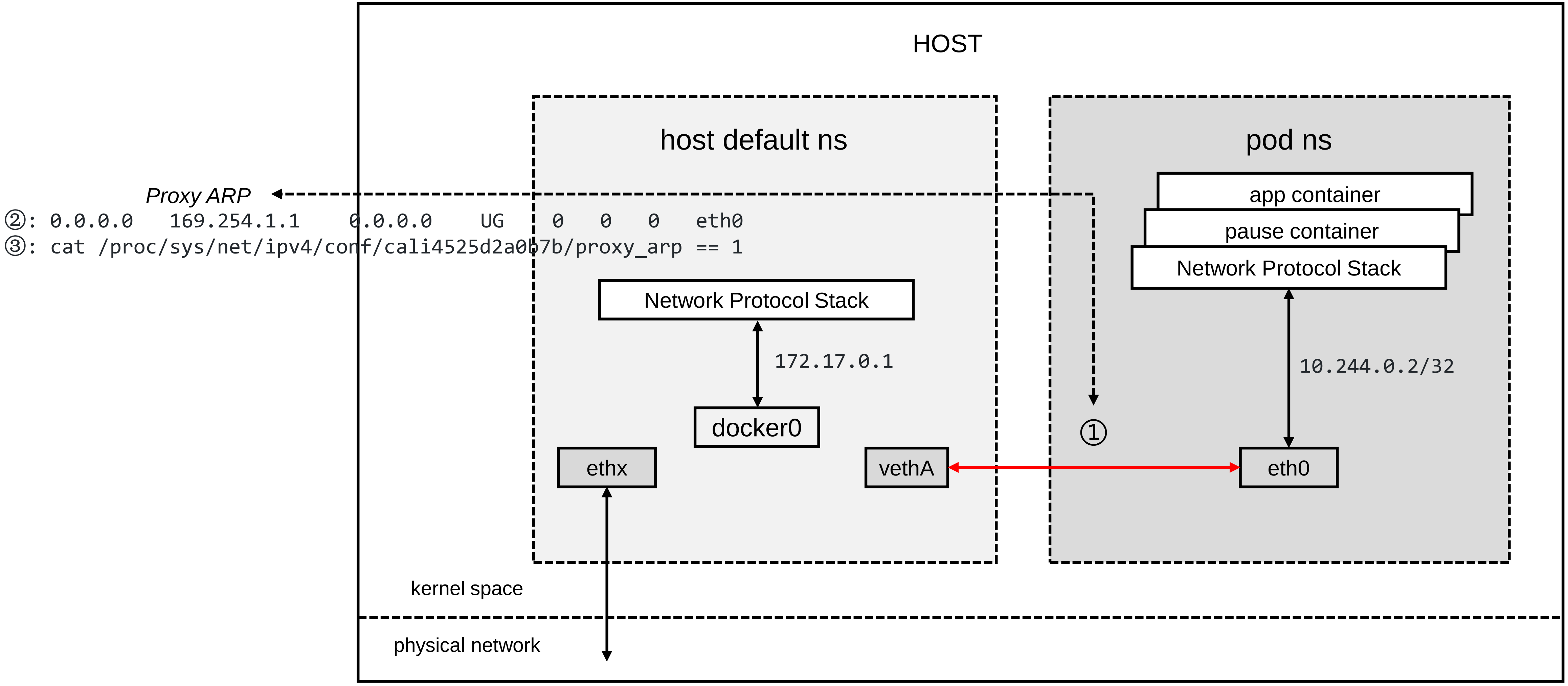
In IP tunnelling, every IP packet, including addressing information of its source and destination IP networks, is encapsulated within another packet format native to the transit network.

At the borders between the source network and the transit network, as well as the transit network and the destination network, gateways are used that establish the end-points of the IP tunnel across the transit network. Thus, the IP tunnel endpoints become native IP routers that establish a standard IP route between the source and destination networks. Packets traversing these end-points from the transit network are stripped from their transit frame format headers and trailers used in the tunnelling protocol and thus converted into native IP format and injected into the IP stack of the tunnel endpoints. In addition, any other protocol encapsulations used during transit, such as IPsec or Transport Layer Security, are removed.

IP in IP, sometimes called ipencap, is an example of IP encapsulation within IP and is described in RFC 2003. Other variants of the IP-in-IP variety are IPv6-in-IPv4 (6in4) and IPv4-in-IPv6 (4in6).

IP tunneling often bypasses simple firewall rules transparently since the specific nature and addressing of the original datagrams are hidden. Content-control software is usually required to block IP tunnels.

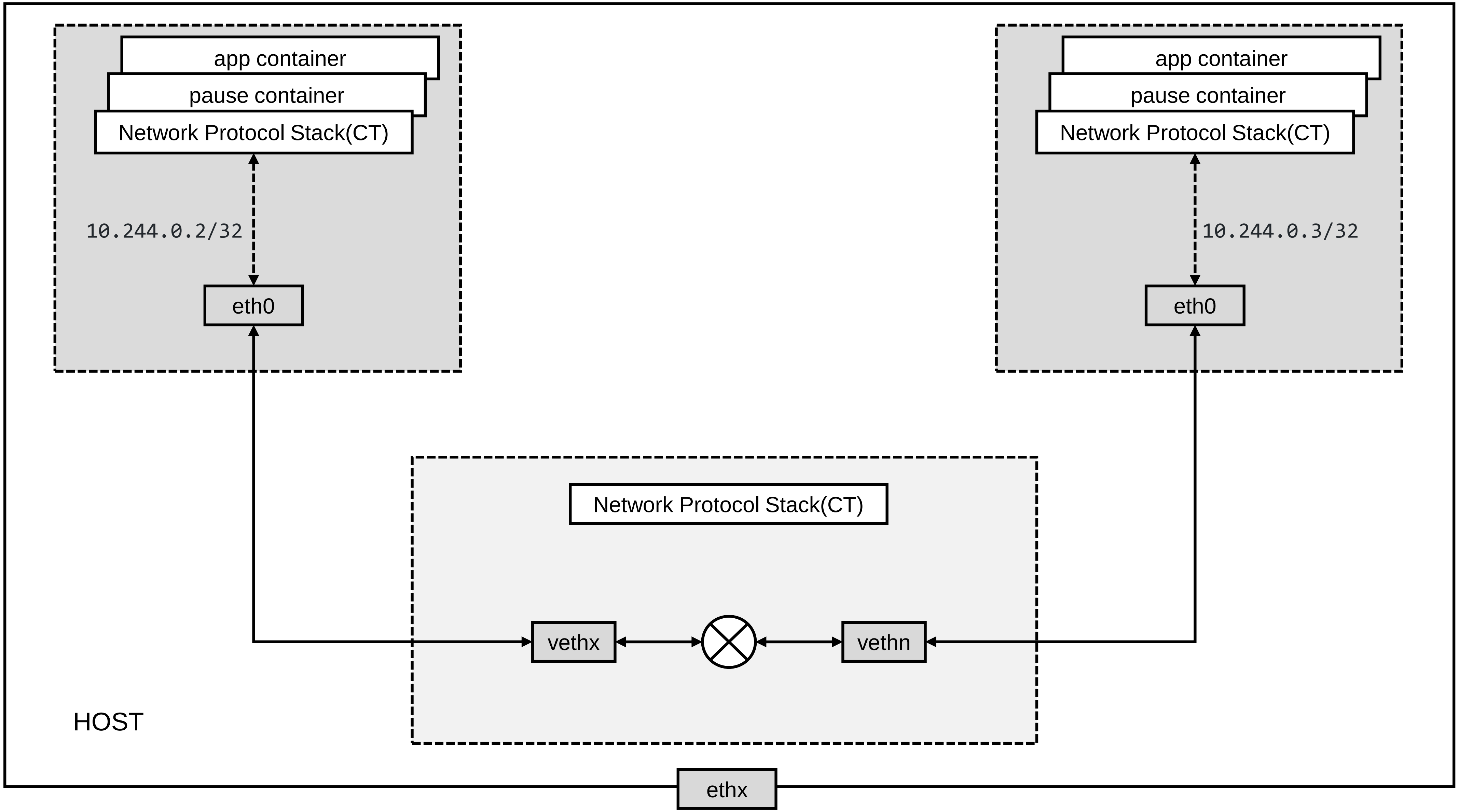
Calico IPIP CrossSubnet Mode - DataPath



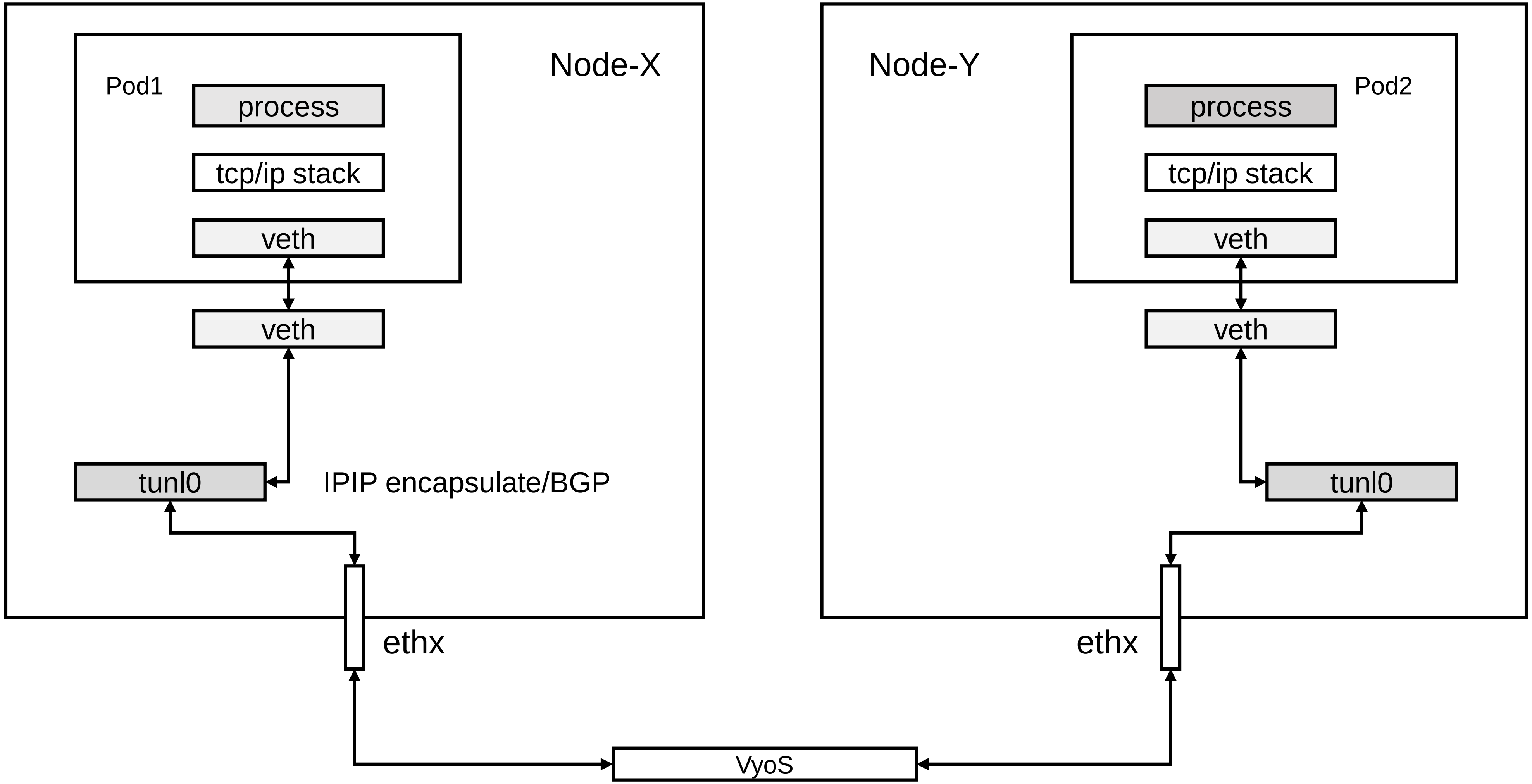
```
②: 0.0.0.0   169.254.1.1    0.0.0.0    UG    0   0   0   eth0
③: cat /proc/sys/net/ipv4/conf/cali4525d2a0b7b/proxy_arp == 1
```

Docker spawns a container in the containers own network namespace (use the CLONE_NEWNET flag defined in sched.h when calling the clone system call to create a new network namespace) and later on runs a veth pair (a cable with two ends) between the container namespace and the host network stack.

Calico IPIP CrossSubnet Mode - DataPath



Calico IPIP CrossSubnet Mode - DataPath



Copyright And Confidentiality

羅偉

牧云/罗伟
Network | IaaS | PaaS | ServiceMesh
交流 学习 沉淀 成长 分享
olaf.luo@foxmail.com
https://www.yuque.com/wei.luo
https://youdianzhishi.com