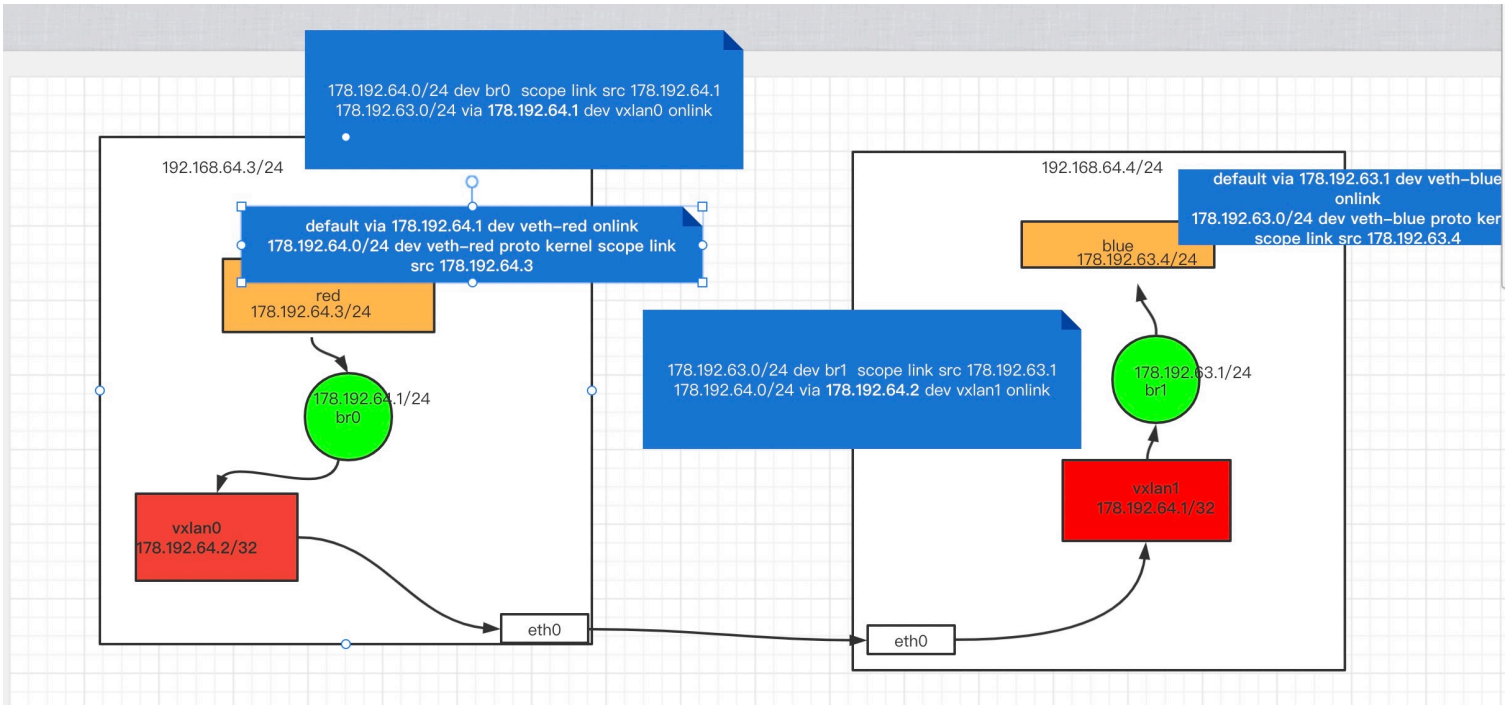
flannel vxlan网络通信原理

vxlan 连通不同主机netns

vxlan这种网络设备，能够对内部要发出去的包进行udp的封装，对外部到达到vxlan上的包进行解包。
vxlan收到包时，会在fdb表中通过要发送包的mac地址去反查要送达的ip地址。

https://www.processon.com/diagraming/61babd535653bb4808a5fe6c




```
## 打开ip forward功能
## 允许防火墙，路由转发
iptables -A FORWARD  -j ACCEPT
## 内核允许路由转发 ,修改值为1
sudo bash -c 'echo 1 > /proc/sys/net/ipv4/ip_forward'


## 创建vxlan interface
 ip link add vxlan0 type vxlan \
    id 42 \
    remote 192.168.64.4 \
    local 192.168.64.3 \
```

```
    dev enp0s1
ip addr add 178.192.64.2/16 dev vxlan0
ip link set vxlan0 up

## 查看fdb表项，vxlan需要查看下一跳的mac地址在对应fdb中的ip地址，让主机上的网卡通过这个ip地址把
bridge fdb
root@primary:/home/ubuntu# bridge fdb
33:33:00:00:00:01 dev enp0s1 self permanent
01:00:5e:00:00:01 dev enp0s1 self permanent
33:33:ff:c0:6d:8a dev enp0s1 self permanent
01:80:c2:00:00:00 dev enp0s1 self permanent
01:80:c2:00:00:03 dev enp0s1 self permanent
01:80:c2:00:00:0e dev enp0s1 self permanent
33:33:00:00:00:01 dev docker0 self permanent
01:00:5e:00:00:6a dev docker0 self permanent
33:33:00:00:00:6a dev docker0 self permanent
01:00:5e:00:00:01 dev docker0 self permanent
02:42:f7:1a:a0:a6 dev docker0 vlan 1 master docker0 permanent
02:42:f7:1a:a0:a6 dev docker0 master docker0 permanent
00:00:00:00:00:00 dev vxlan0 dst 192.168.64.4 via enp0s1 self permanent

## 创建命名空间
ip netns add  red
sudo ip link add veth-red type veth peer name veth-red-br
ip link set veth-red up
ip link set veth-red-br up
## 创建网桥
brctl addbr br0
ip link set br0 up
sudo ip link set veth-red netns red
sudo ip link set veth-red-br master br0

## 绑定ip
ip netns exec red ip addr add 178.192.64.3/24 dev veth-red
ip addr add  178.192.64.1/24 dev br0

## 创建路由
ip netns exec red ip route add default via 178.192.64.1  dev veth-red onlink
```

flannel 在k8s上不同网段通信

k8s上的vxlan配置方式和上面所讲的略有不同

不同点

之前fdb表配置的映射为

```
00:00:00:00:00:00 dev vxlan0 dst 192.168.64.4 via enp0s1 self permanent
```

这样存在一个问题，任何封装的vxlan包都将发往192.168.64.4主机，如果集群是多台主机的情况呢，则不能这么配置。
所以flannel明确配置了要发往192.168.64.4 主机的包的mac地址为该主机上的vxlan设备的mac地址。
既然明确配置了mac地址，回到配置路由的地方.

```
178.192.63.0/24 via 178.192.64.1 dev vxlan0 onlink
```

这里从vxlan0设备出去的下一跳地址是178.192.64.1，那么如何让178.192.64.1 和刚才所讲的mac地址建立映射呢?
所以flannel又配置了arp表，这样就能让访问特定ip段的包送到特定的主机上，由目的主机的vxlan包进行解包。

```
## nud permanent指明该arp记录不会过期，不用做存活检查
ip neigh add 178.192.64.1 lladdr 02:3f:39:67:7d:f9 dev vxlan0 nud permanent
```

总结

flannel 所做的操作就是在节点新增时，为其他节点动态的创建路由，动态的更新arp，fdb表，从而让不同主机上的网络包能互相传递。