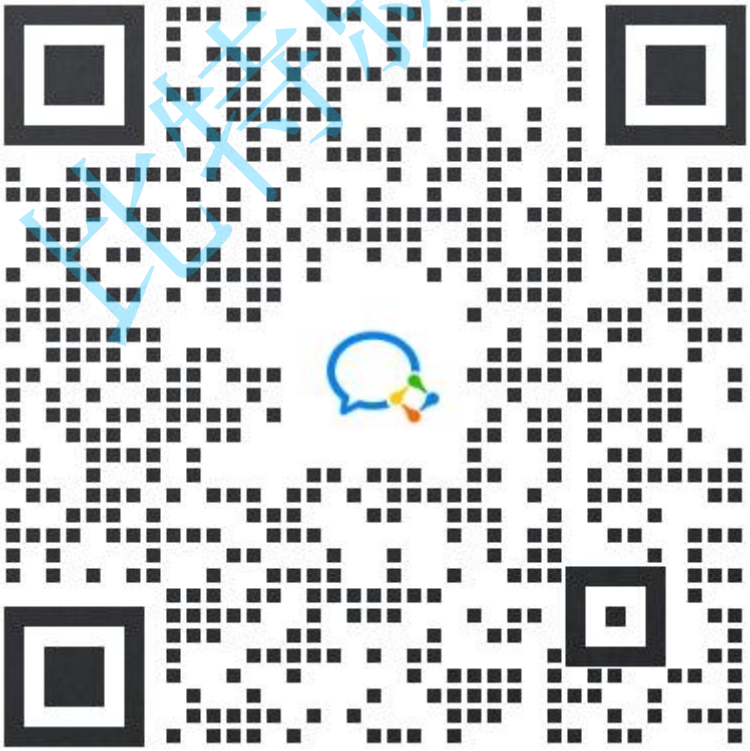
网桥实战

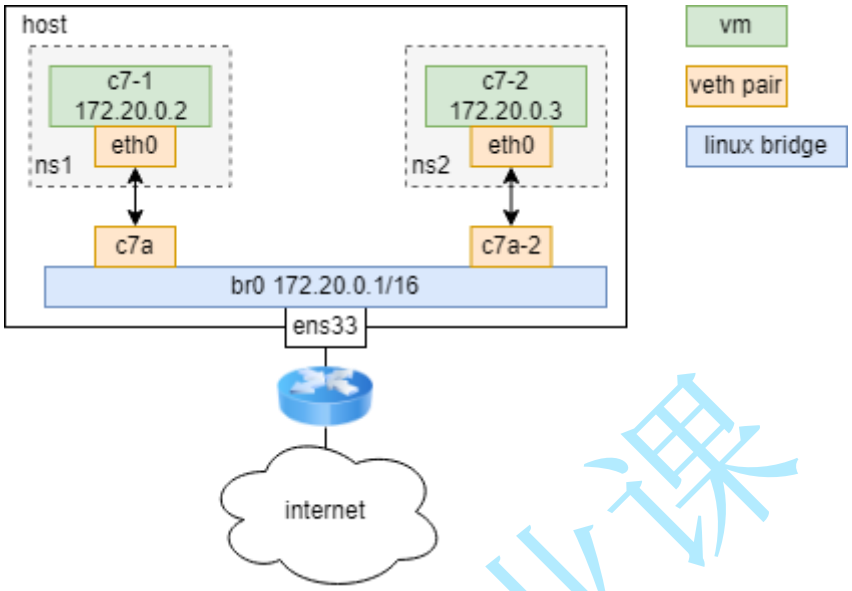
版权说明

对比特项目感兴趣，可以联系这个微信。



基础知识

Linux Bridge（网桥）是用纯软件实现的虚拟交换机，有着和物理交换机相同的功能，例如二层交换，MAC 地址学习等。因此我们可以把 tun/tap，veth pair 等设备绑定到网桥上，就像是把设备连接到物理交换机上一样。此外它和 veth pair、tun/tap 一样，也是一种虚拟网络设备，具有虚拟设备的所有特性，例如配置 IP，MAC 地址等。



linux 提供了 brctl 工具来管理和查看网桥

1. 安装方式

```Shell
# centos
yum install -y bridge-utils
# ubuntu
apt-get install -y bridge-utils
```

2. 新建一个网桥:

```Plain Text
brctl addbr <bridge>
```

3. 添加一个设备（例如 `eth0`）到网桥:

```Plain Text
brctl addif <bridge> eth0
```

4. 显示当前存在的网桥及其所连接的网络端口:

```Plain Text
brctl show
```

5. 启动网桥:

```
Plain Text
ip link set <bridge> up
```

6. 删除网桥，需要先关闭它:

```
Plain Text
ip link set <bridge> down
brctl delbr <bridge>
```

或者使用 `ip link del` 命令直接删除网桥

```
Plain Text
ip link del <bridge>
```

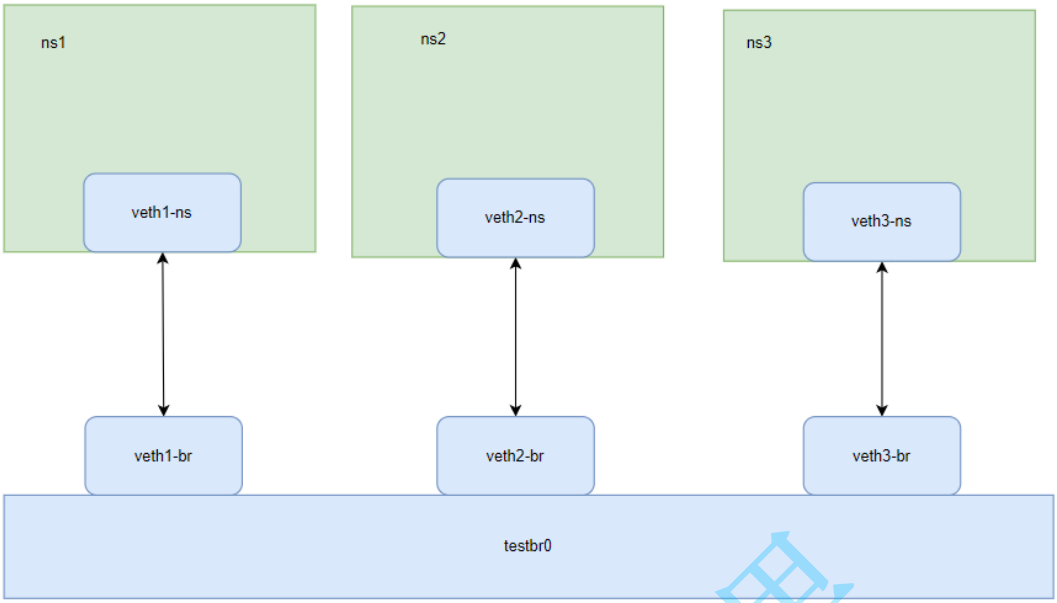
> 增加 Linux Bridge 时会自动增加一个同名虚拟网卡在宿主机器上，因此我们可以通过 `ip link` 命令操作这个虚拟网卡，实际上也就是操作网桥，并且只有当这个虚拟网卡状态处于 up 的时候，网桥才会转发数据。

实战目的

了解网桥的搭建和作用。

实战步骤

1. 我们创建三个 netns，三对 veth pair，分别一端在 netns 中，另一端连接在网桥上，网络拓扑如下



2. 新建网络命名空间

```Shell
root@139-159-150-152:~# ip netns add ns1
root@139-159-150-152:~# ip netns add ns2
root@139-159-150-152:~# ip netns add ns3
```

3. 创建 veth 对

```Shell
root@139-159-150-152:~# ip link add veth1-ns type veth peer name
veth1-br
root@139-159-150-152:~# ip link add veth2-ns type veth peer name
veth2-br
root@139-159-150-152:~# ip link add veth3-ns type veth peer name
veth3-br
```

4. 将 ns 一段的网卡移入到命名空间

```Shell
root@139-159-150-152:~# ip link set dev veth1-ns netns ns1
root@139-159-150-152:~# ip link set dev veth2-ns netns ns2
root@139-159-150-152:~# ip link set dev veth3-ns netns ns3
```

5. 启动网卡，并配置 ip，开启本地回环，可以 ping 自己

```
Shell
root@139-159-150-152:~# ip netns exec ns1 ip link set veth1-ns up
root@139-159-150-152:~# ip netns exec ns2 ip link set veth2-ns up
root@139-159-150-152:~# ip netns exec ns3 ip link set veth3-ns up

root@139-159-150-152:~# ip netns exec ns1 ip link set lo up
root@139-159-150-152:~# ip netns exec ns2 ip link set lo up
root@139-159-150-152:~# ip netns exec ns3 ip link set lo up

root@139-159-150-152:~# ip netns exec ns1 ip addr add
10.100.0.11/24 dev veth1-ns
root@139-159-150-152:~# ip netns exec ns2 ip addr add
10.100.0.12/24 dev veth2-ns
root@139-159-150-152:~# ip netns exec ns3 ip addr add
10.100.0.23/24 dev veth3-ns
```

6. 测试网络联通性，此时是不通的

```
Shell
root@139-159-150-152:~# ip netns exec ns3 ping 10.100.0.11
PING 10.100.0.11 (10.100.0.11) 56(84) bytes of data.
^C
--- 10.100.0.11 ping statistics ---
9 packets transmitted, 0 received, 100% packet loss, time 8174ms
```

7. 创建网桥

```
Shell
root@139-159-150-152:~# brctl addbr testbr0
```

8. 查看 ifconfig 可以看到网桥也像网卡一样可以配置 ip 信息

```
Shell
root@139-159-150-152:~# ifconfig
br-df863876204e: flags=4099<UP,BROADCAST,MULTICAST>  mtu 1500
        inet 192.168.0.1  netmask 255.255.0.0  broadcast
192.168.255.255
        inet6 fe80::42:20ff:feb7:55bb  prefixlen 64  scopeid
0x20<link>
        ether 02:42:20:b7:55:bb  txqueuelen 0  (Ethernet)
```

```
        RX packets 0  bytes 0 (0.0 B)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 29  bytes 4216 (4.2 KB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

docker0: flags=4099<UP,BROADCAST,MULTICAST>  mtu 1500
        inet 172.17.0.1  netmask 255.255.0.0  broadcast
172.17.255.255
        inet6 fe80::42:d7ff:fe87:d11b  prefixlen 64  scopeid
0x20<link>
        ether 02:42:d7:87:d1:1b  txqueuelen 0  (Ethernet)
        RX packets 586466  bytes 36737314 (36.7 MB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 860320  bytes 1655179375 (1.6 GB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.0.183  netmask 255.255.255.0  broadcast
192.168.0.255
        inet6 fe80::f816:3eff:fe9d:f4ac  prefixlen 64  scopeid
0x20<link>
        ether fa:16:3e:9d:f4:ac  txqueuelen 1000  (Ethernet)
        RX packets 4036983  bytes 4360556431 (4.3 GB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 2646870  bytes 470938993 (470.9 MB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING>  mtu 65536
        inet 127.0.0.1  netmask 255.0.0.0
        inet6 ::1  prefixlen 128  scopeid 0x10<host>
        loop  txqueuelen 1000  (Local Loopback)
        RX packets 1126921  bytes 449483727 (449.4 MB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 1126921  bytes 449483727 (449.4 MB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

lxcbr0: flags=4099<UP,BROADCAST,MULTICAST>  mtu 1500
        inet 10.0.3.1  netmask 255.255.255.0  broadcast 10.0.3.255
        inet6 fe80::216:3eff:fe00:0  prefixlen 64  scopeid
0x20<link>
        ether 00:16:3e:00:00:00  txqueuelen 1000  (Ethernet)
        RX packets 595  bytes 72213 (72.2 KB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 817  bytes 73558 (73.5 KB)
```

```
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

root@139-159-150-152:~# ifconfig -a
br-df863876204e: flags=4099<UP,BROADCAST,MULTICAST>  mtu 1500
        inet 192.168.0.1  netmask 255.255.0.0  broadcast
192.168.255.255
        inet6 fe80::42:20ff:feb7:55bb  prefixlen 64  scopeid
0x20<link>
        ether 02:42:20:b7:55:bb  txqueuelen 0  (Ethernet)
        RX packets 0  bytes 0 (0.0 B)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 29  bytes 4216 (4.2 KB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

docker0: flags=4099<UP,BROADCAST,MULTICAST>  mtu 1500
        inet 172.17.0.1  netmask 255.255.0.0  broadcast
172.17.255.255
        inet6 fe80::42:d7ff:fe87:d11b  prefixlen 64  scopeid
0x20<link>
        ether 02:42:d7:87:d1:1b  txqueuelen 0  (Ethernet)
        RX packets 586466  bytes 36737314 (36.7 MB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 860320  bytes 1655179375 (1.6 GB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

eth0: flags=4163<UP,BROADCAST,RUNNING,MULTICAST>  mtu 1500
        inet 192.168.0.183  netmask 255.255.255.0  broadcast
192.168.0.255
        inet6 fe80::f816:3eff:fe9d:f4ac  prefixlen 64  scopeid
0x20<link>
        ether fa:16:3e:9d:f4:ac  txqueuelen 1000  (Ethernet)
        RX packets 4036994  bytes 4360557213 (4.3 GB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 2646885  bytes 470945451 (470.9 MB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

lo: flags=73<UP,LOOPBACK,RUNNING>  mtu 65536
        inet 127.0.0.1  netmask 255.0.0.0
        inet6 ::1  prefixlen 128  scopeid 0x10<host>
        loop  txqueuelen 1000  (Local Loopback)
        RX packets 1126921  bytes 449483727 (449.4 MB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 1126921  bytes 449483727 (449.4 MB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

```
lxcbr0: flags=4099<UP,BROADCAST,MULTICAST>  mtu 1500
        inet 10.0.3.1  netmask 255.255.255.0  broadcast 10.0.3.255
        inet6 fe80::216:3eff:fe00:0  prefixlen 64  scopeid
0x20<link>
        ether 00:16:3e:00:00:00  txqueuelen 1000  (Ethernet)
        RX packets 595  bytes 72213 (72.2 KB)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 817  bytes 73558 (73.5 KB)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

testbr0: flags=4098<BROADCAST,MULTICAST>  mtu 1500
        ether 4e:9e:0e:a6:57:2d  txqueuelen 1000  (Ethernet)
        RX packets 0  bytes 0 (0.0 B)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 0  bytes 0 (0.0 B)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

veth1-br: flags=4098<BROADCAST,MULTICAST>  mtu 1500
        ether f2:8f:b4:5c:0f:cf  txqueuelen 1000  (Ethernet)
        RX packets 0  bytes 0 (0.0 B)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 0  bytes 0 (0.0 B)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

veth2-br: flags=4098<BROADCAST,MULTICAST>  mtu 1500
        ether 06:ab:bb:e0:c4:1f  txqueuelen 1000  (Ethernet)
        RX packets 0  bytes 0 (0.0 B)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 0  bytes 0 (0.0 B)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0

veth3-br: flags=4098<BROADCAST,MULTICAST>  mtu 1500
        ether aa:16:a5:1d:b1:5a  txqueuelen 1000  (Ethernet)
        RX packets 0  bytes 0 (0.0 B)
        RX errors 0  dropped 0  overruns 0  frame 0
        TX packets 0  bytes 0 (0.0 B)
        TX errors 0  dropped 0 overruns 0  carrier 0  collisions 0
```

9. 启动网桥

```
Shell
root@139-159-150-152:~# ip link set testbr0 up
```

10. 给网桥设置 ip 地址

```
Shell
root@139-159-150-152:~# ip addr add 10.100.0.1/24 dev testbr0
```

11. 启动连接到网桥的网卡

```
Shell
root@139-159-150-152:~# ip link set veth1-br up
root@139-159-150-152:~# ip link set veth2-br up
root@139-159-150-152:~# ip link set veth3-br up
```

12. 将另一端的网卡插上网桥

```
Shell
root@139-159-150-152:~# brctl addif testbr0 veth1-br
root@139-159-150-152:~# brctl addif testbr0 veth2-br
root@139-159-150-152:~# brctl addif testbr0 veth3-br
```

13. 测试网络联通性，此时是可以连通的

```
Shell
root@139-159-150-152:~# ip netns exec ns3 ping 10.100.0.11
PING 10.100.0.11 (10.100.0.11) 56(84) bytes of data.
64 bytes from 10.100.0.11: icmp_seq=1 ttl=64 time=0.072 ms
64 bytes from 10.100.0.11: icmp_seq=2 ttl=64 time=0.036 ms
^C
```

14. 如果发现不能转发，原因是 linux 加入了 bridge_netfilter。需要开启允许通过

```
Shell
 iptables -A FORWARD -i testbr0 -j ACCEPT
```

15. 清理空间

```
Shell
ip link del veth1-br
ip link del veth2-br
```

```
ip link del veth3-br
ip link del testbr0
ip netns del ns1
ip netns del ns2
ip netns del ns3

#iptables 规则清理!!!!! 慎重操作，不要删除错 id，可能会导致各种网络问题，不影响系统使用也可以不操作
root@139-159-150-152:~# iptables --line-numbers -nvL
Chain INPUT (policy ACCEPT 124 packets, 12696 bytes)
num   pkts bytes target     prot opt in     out     source
destination
1        0     0 ACCEPT     tcp  --  lxcbr0 *       0.0.0.0/0
0.0.0.0/0            tcp dpt:53
2        5   368 ACCEPT     udp  --  lxcbr0 *       0.0.0.0/0
0.0.0.0/0            udp dpt:53
3        0     0 ACCEPT     tcp  --  lxcbr0 *       0.0.0.0/0
0.0.0.0/0            tcp dpt:67
4        4  1265 ACCEPT     udp  --  lxcbr0 *       0.0.0.0/0
0.0.0.0/0            udp dpt:67

Chain FORWARD (policy DROP 0 packets, 0 bytes)
num   pkts bytes target     prot opt in     out     source
destination
1       26  2184 ACCEPT     all  --  *      lxcbr0  0.0.0.0/0
0.0.0.0/0
2       26  2184 ACCEPT     all  --  lxcbr0 *       0.0.0.0/0
0.0.0.0/0
3      147 11696 DOCKER-USER  all  --  *      *       0.0.0.0/0
0.0.0.0/0
4      147 11696 DOCKER-ISOLATION-STAGE-1  all  --  *      *
0.0.0.0/0            0.0.0.0/0
5        0     0 ACCEPT     all  --  *      br-df863876204e
0.0.0.0/0            0.0.0.0/0            ctstate
RELATED,ESTABLISHED
6        0     0 DOCKER     all  --  *      br-df863876204e
0.0.0.0/0            0.0.0.0/0
7        0     0 ACCEPT     all  --  br-df863876204e !br-
df863876204e  0.0.0.0/0            0.0.0.0/0
8        0     0 ACCEPT     all  --  br-df863876204e br-
df863876204e  0.0.0.0/0            0.0.0.0/0
9      860K 1643M ACCEPT     all  --  *      docker0  0.0.0.0/0
0.0.0.0/0            ctstate RELATED,ESTABLISHED
10     128  6441 DOCKER     all  --  *      docker0  0.0.0.0/0
```

```
0.0.0.0/0
11     586K   37M ACCEPT     all  --  docker0 !docker0  0.0.0.0/0
0.0.0.0/0
12       1    84 ACCEPT     all  --  docker0 docker0  0.0.0.0/0
0.0.0.0/0
13       0     0 ACCEPT     all  --  testbr0 *       0.0.0.0/0
0.0.0.0/0

#删除规则，慎重不要删除错了
root@139-159-150-152:~# iptables -D FORWARD 13
root@139-159-150-152:~# iptables --line-numbers -nvL |grep test
root@139-159-150-152:~#
```