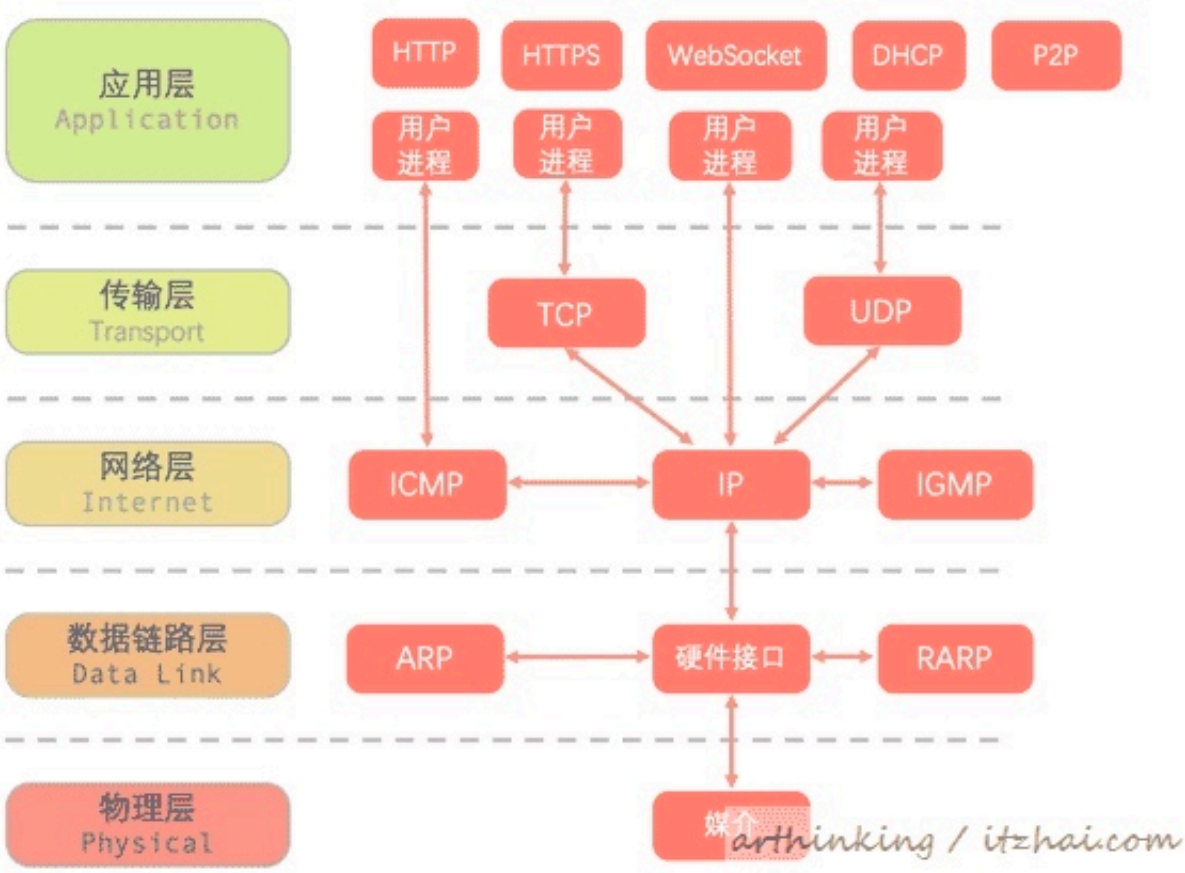
包流转路径

发包流程，一个ip网络包发送前，需要通过arp协议去获取mac地址（如果本地没有arp缓存的话），获取到mac地址后，内核会封装网络ip包对网络包进行发送。

包如何在局域网传播，如何传递到其他网络。
图像来自:https://www.cnblogs.com/arthinking/p/13429848.html



几个规则：

如何选择”下一跳“(nexthop)

网络数据包是一台台主机接力传递的，到达目的地后，又是一台台主机接力返回回来。
那么如何选择”下一跳“(nexthop)的主机。
1, 在发送ip包之前，会把发送的目的ip与本地路由表去进行比较，看ip包是从哪个网卡出去。

通过ip去寻找路由规则rule，通过路由规则rule去特定的路由表去寻找路由策略。

ip rule

字段	含义
XX	第一列数字是优先级，小的数字优先级高
lookup [xxx]	表示搜索xxx路由表，1-252之间的数字或名称
scope link	为要发往的目的网段

```
## 系统默认有3条记录
0: from all lookup local
32766: from all lookup main
32767: from all lookup default
```

ip route

字段	含义
10.211.55.0/24	为要发往的目的网段
dev eth0 proto kernel	由网卡eth0发送，eth0是由内核安装的
src 10.211.55.6	发送出去的ip地址为 10.211.55.6

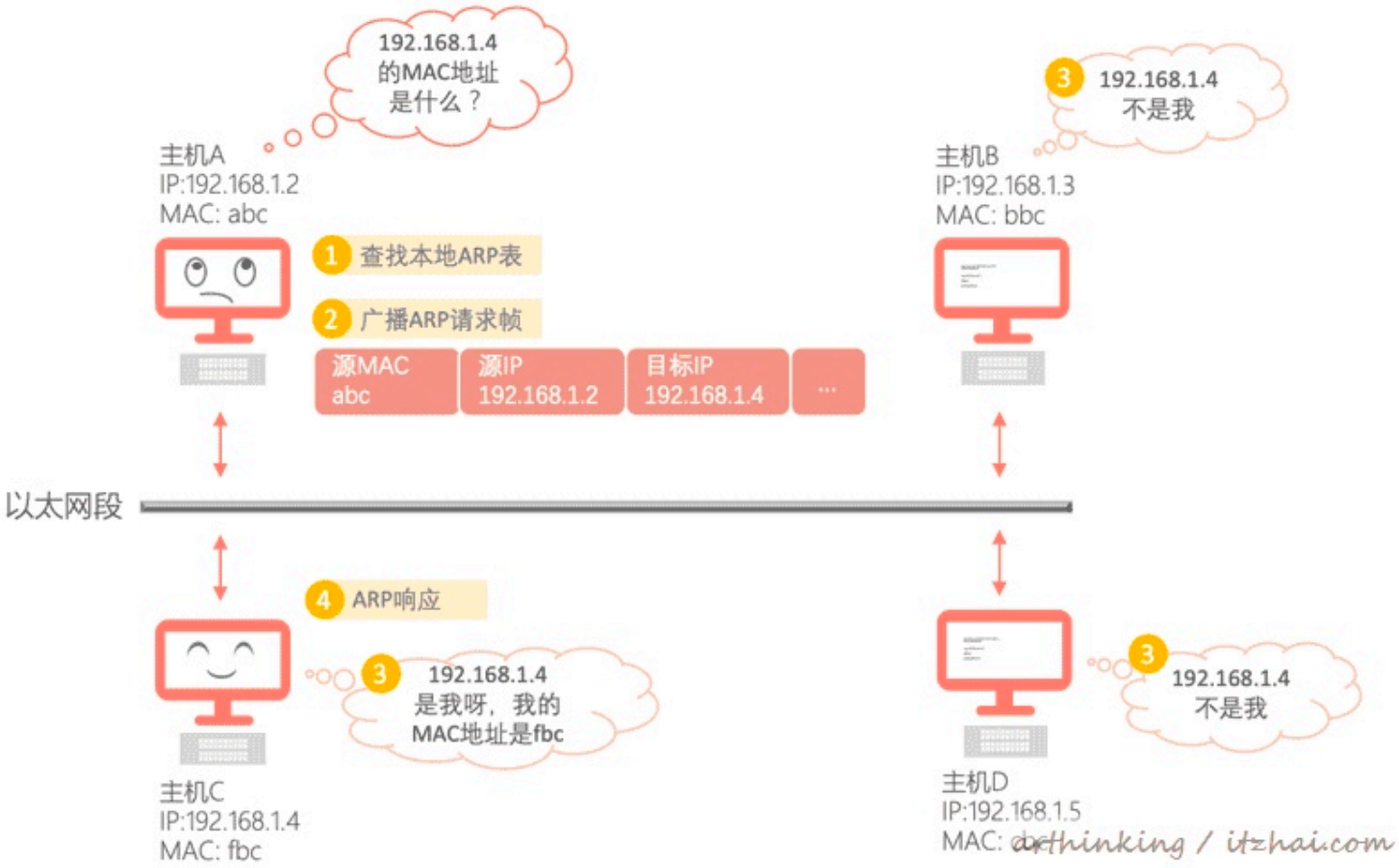
```
$ ip route
default via 10.211.55.1 dev eth0 proto dhcp src 10.211.55.6 metric 100
## 发往 10.211.55.0/24 这个网段的包会经由 网卡eth0 发送出去 ip为10.211.55.6 10.211.55.0/
10.211.55.1 dev eth0 proto dhcp scope link src 10.211.55.6 metric 100
172.17.0.0/16 dev docker0 proto kernel scope link src 172.17.0.1
```

2，如果本地都没有路由，则默认发往网关的地址，不论是发往本地的路由还是网关路由，发送之前都会通过arp协议获取网关或者对应路由网卡的mac地址 。

3， mac地址得到以后，就将数据包发往nexthop地点。

arp协议

arp 发的链路层的包不能跨网段。
在一个局域网里，通常会存在交换机，交换机的作用是将局域网里的机器通过网线连接起来，当一个数据包到达交换机的时候，交换机会通过包的目的mac地址看是单播还是广播。只要连上了端口就能发出去。



路由器

路由器转发的是ip网络包，路由器一般也具有交换机的功能。
之前提到如果本地没有对应网络包的路由就会将包发往默认网关，一般配置的默认网关ip所在的机

器具有路由功能。将ip网络包收到以后，如果目的ip不是在接收网卡的网段，路由器就会去寻找自己机器上其他网卡，看是否有相同网段的网卡，有的话就从那个网卡把包发出去，否则就走它的默认网关出去。

namespace

内核的功能，能够让每个网络名字空间都有独立的网络配置，比如：网络设备、路由表，ARP表等。

```
## 查看网络命名空间
ip netns
## 添加网络命名空间
ip netns add 空间名
```

bridge

bridge 是linux上的交换机，linux开启ip forward功能后，并赋予bridge 一个ip能让bridge具有路由功能，转发不同网段的网络包。

```
## 查看网桥
brctl show
## 添加网桥
brctl addbr 网桥名
```

veth-pair

linux上的一个网络设备，这个网络设备有两个端点，数据从一个端点进入，必然从另外一个端点流出。每个veth都可以被赋予IP地址，并参与三层网络路由过程。

网络设备被赋予ip之后，那它的一端就可以被认为与协议栈相连。

```
ip netns add test_veth
ip link add veth-test type veth peer name veth-test2
ip link set veth-test2 netns test_veth
ip link set veth-test up
ip link set veth-test2 up
ip link addr add 10.11.13.3/24 dev veth-test
ip link addr add 10.11.14.4/24 dev veth-test2
## 添加路由，让访问10.11.14.0/24网段的包从veth-test出去
```

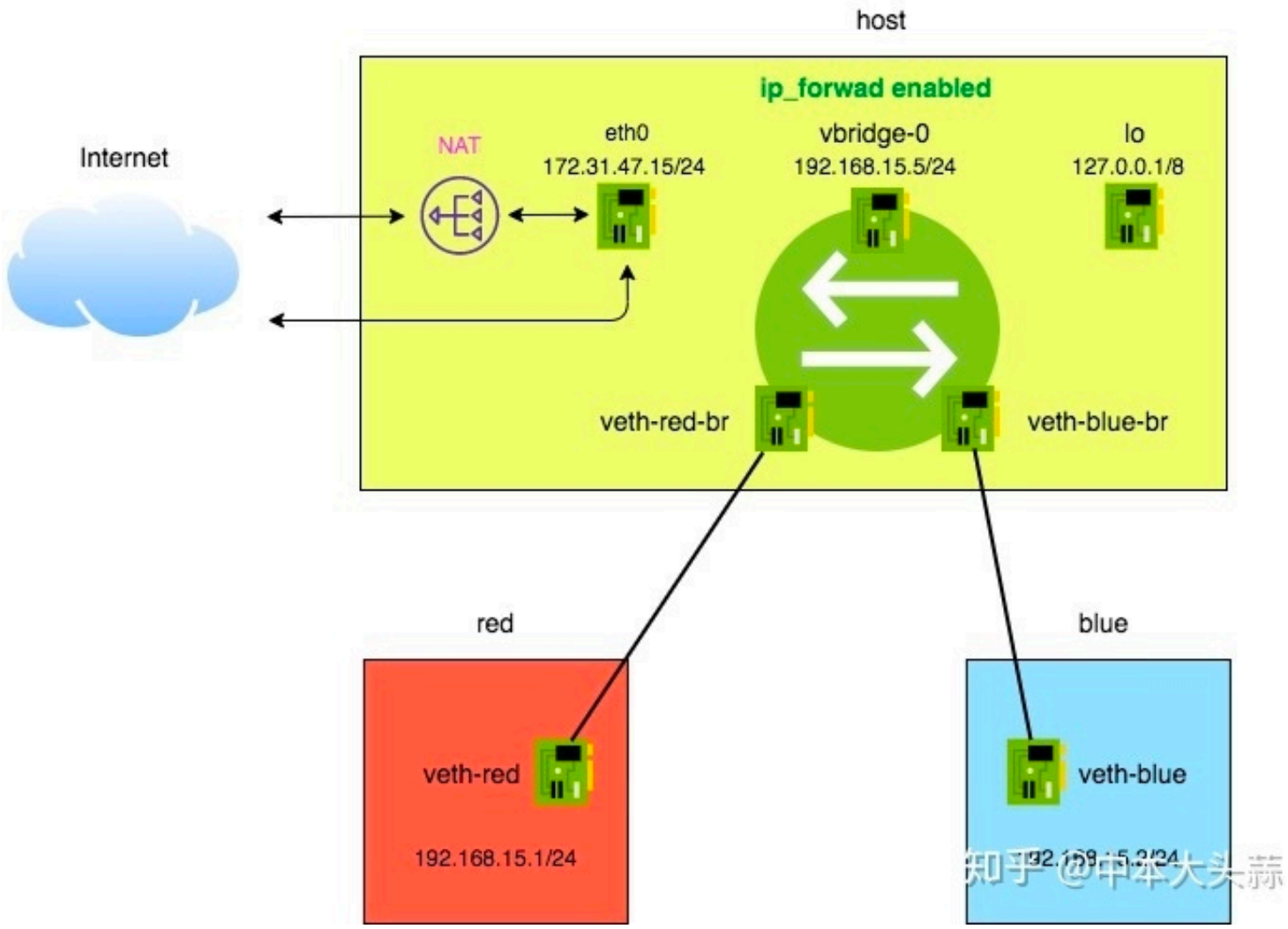
```
ip route add 10.11.14.0/24 dev veth-test

## test_veth 命名空间添加默认路由
ip route add default dev veth-test2
```

docker里的网络拓扑结构

3个问题

- 容器之间通信
- 容器访问外网
- 公网访问容器

图像来自:https://zhuanlan.zhihu.com/p/199298498



前置条件:

```
## 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'
```

```
## 创建网校
brctl addbr br0
## 创建命名空间
sudo ip netns  add red
## 创建veth pair
sudo ip link add veth-red type veth peer name veth-red-br
sudo ip link add veth-blue type veth peer name veth-blue-br
## 将veth-red放到red namespace里
sudo ip link set veth-red netns red
sudo ip link set veth-blue netns blue
## veth一端链接到网桥
sudo ip link set veth-red-br master br0
sudo ip link set veth-blue-br master br0


## 为网桥赋上ip
ip addr add 192.168.15.5/24 dev br0


## 防火墙nat设置
iptables -t nat -A POSTROUTING -s 192.168.15.0/24 -j MASQUERADE
```