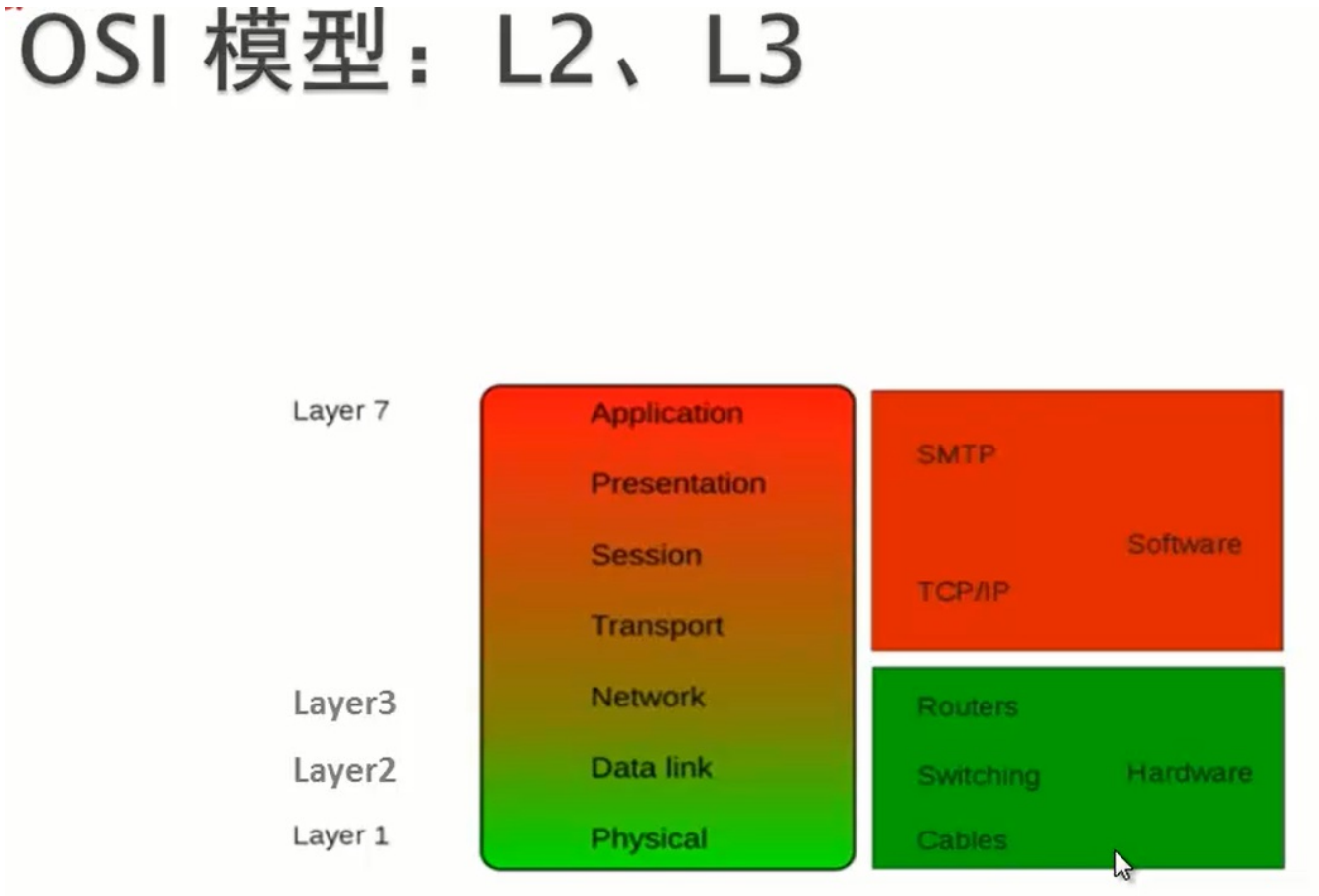
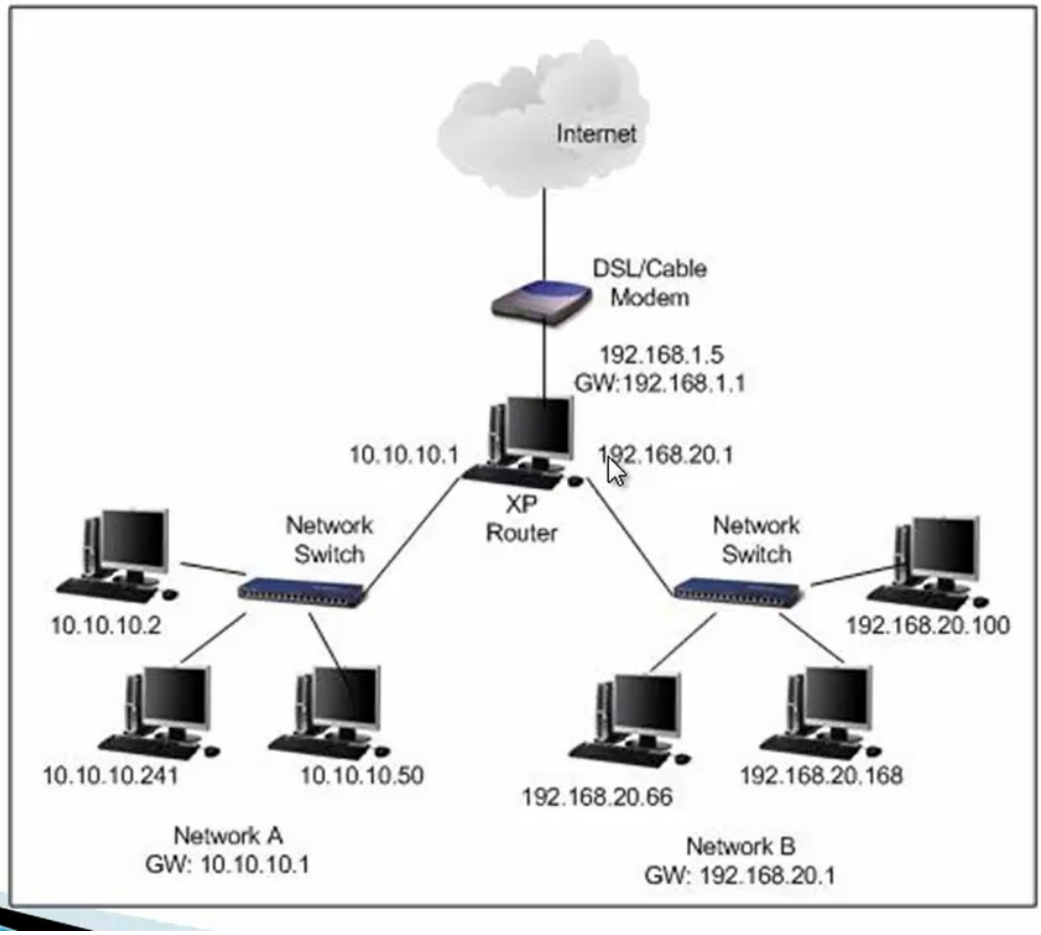

OSI 模型：L2、L3
Layer 7
Application
Presentation
Session
Transport
Network
Data link
Physical
Layer3
Layer2
Layer 1
SMTP
Software
TCP/IP
Routers
Switching
Hardware
Cables

交换机、路由器



区别有哪些

- 工作层次不同
  - L2/L3
- 数据转发依据对象不同
  - 数据帧（MAC）/IP数据包（IP）
- 解决问题不同
  - 同网段互通/多网段互通（路由）

路由表

Routing Table

```
[root@network0 ~]# route
Kernel IP routing table
Destination     Gateway         Genmask         Flags Metric Ref    Use Iface
10.20.0.0       *               255.255.255.0   U     0      0        0 eth0
192.168.4.0     *               255.255.255.0   U     0      0        0 eth2
172.16.0.0      *               255.255.255.0   U     0      0        0 br-ex
link-local      *               255.255.0.0     U     1002   0        0 eth0
link-local      *               255.255.0.0     U     1003   0        0 eth1
link-local      *               255.255.0.0     U     1004   0        0 eth2
default         10.20.0.1       0.0.0.0         UG    0      0        0 eth0
[root@network0 ~]#
```

比如去往10.20.0.0这个地址的数据包经过本机则通过eth0网口出去，其实就是一些路由规则的设定

IP Table

其实IP Table和路由表都是数据域三层上面的一些功能

IP Table

```
[root@vm1 ~]# ip netns exec router-ns iptables -t nat -nL
Chain PREROUTING (policy ACCEPT)
target     prot opt source               destination
DNAT       all  --  0.0.0.0/0            192.168.4.51        to:192.168.1.11

Chain POSTROUTING (policy ACCEPT)
target     prot opt source               destination
SNAT       all  --  192.168.1.11         0.0.0.0/0           to:192.168.4.51
SNAT       all  --  192.168.1.0/24       0.0.0.0/0           to:192.168.4.50

Chain OUTPUT (policy ACCEPT)
target     prot opt source               destination
DNAT       all  --  0.0.0.0/0            192.168.4.51        to:192.168.1.11
[root@vm1 ~]#
```

三种链条：

PREROUTING：把发往192.168.4.51的目的地址替换成192.168.1.11，这里做了一个转发。

DHCP（动态主机配置协议）

- 功能
  - 统一网络主机分配IP地址
- 好处
  - 降低了配置和部署设备时间。
  - 降低了发生配置错误的可能性。
  - 可以集中化管理设备的IP地址分配。

Linux 实现

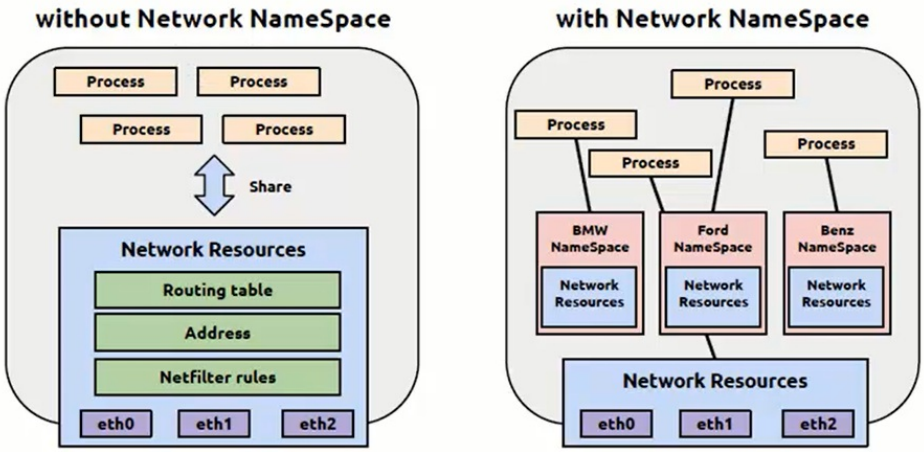
- 工具软件
  - dnsmasq

```
[root@network0 ~]# ps -ef|grep dns
root     13396 11985  0 23:26 pts/1    00:00:00 grep dns
nobody   23523     1  0 Aug16 ?        00:00:00 dnsmasq --no-hosts --no-resolv --strict-order
--bind-interfaces --interface=tap165eb9ab-dd --except-interface=lo --pid-file=/var/lib/neutron
/dhcp/1ddab5f1-c411-4f08-8e65-edb582309a4c/pid --dhcp-hostsfile=/var/lib/neutron/dhcp/1ddab5f1
-c411-4f08-8e65-edb582309a4c/host --addn-hosts=/var/lib/neutron/dhcp/1ddab5f1-c411-4f08-8e65-e
db582309a4c/addn_hosts --dhcp-optsfile=/var/lib/neutron/dhcp/1ddab5f1-c411-4f08-8e65-edb582309
a4c/opts --leasefile-ro --dhcp-range=tag0,192.168.1.0,static,86400s --dhcp-lease-max=256 --con
f-file= --domain=openstacklocal
[root@network0 ~]#
```

这个dhcp服务作用在上述interface网卡上，跟这个网卡相连的都被它分配

网络命名空间

- 独享网络资源
  - Interface
  - Iptables
  - Router
- LXC
  - 网络隔离
  - 网络 overlay



放在不同命名空间的网络资源是互不可见的
interface、iptables表、路由表分离开
一般用来网络隔离，同时可以配置。当然进程是可见的但是网络资源是不可见的。

叠加网络（Network Overlay）

1. 一个数据包(或帧)封装在另一个数据包内;被封装的包转发到隧道端点后再被拆装。

2. 叠加网络就是使用这种所谓“包内之包”的技术安全地将一个网络隐藏在另一个网络中，然后将网络区段进行迁移。

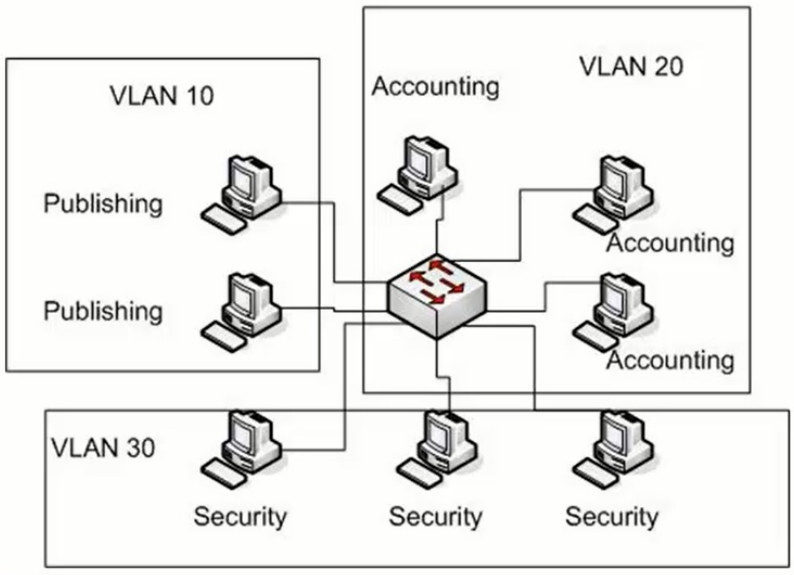
- Vlan
  - L2 over L2
- GRE
  - L3 over L3（UDP）
- Vxlan
  - L2 over L3（UDP）

解决问题

- 数据中心网络数量限制
  - 1-> 4096 -> 1600万
- 物理网络基础设施限制
  - 不改变物理网络变更VM网络拓扑
  - VM迁移
- 多租户场景
  - 支持IP地址重叠

Vlan （虚拟局域网）

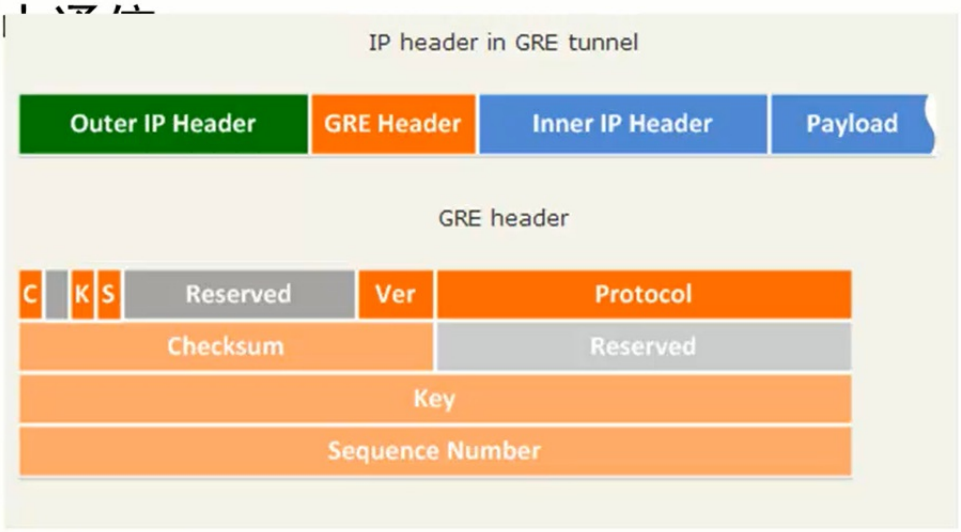
- 2层广播隔离
- 灵活的组网，IP地址划分
- 最多4096
- 直接在L2实现



有Vlan tag，总共有4096个网络。也是2层数据包，只是带上了vlan tag而已

通用路由封装协议(GRE)

- 夸不同网络实现二次IP通信
- L3上面包装L3
- 封装在IP报文中
- 点对点隧道



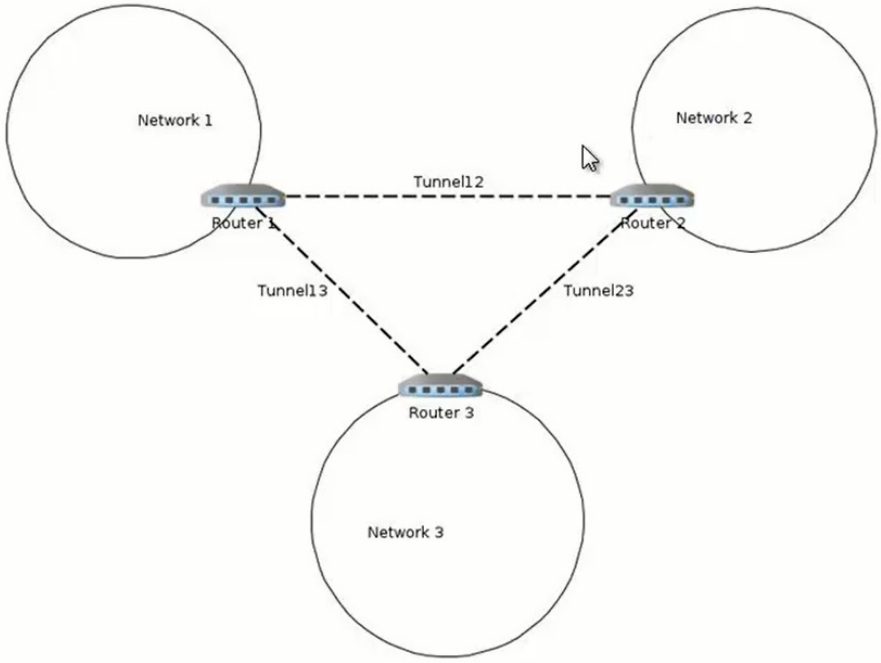
利用点对点隧道达到不同网络直接的通信.

GRE用在SDN中的好处

- 不用变更底层网络架构重建L2、L3通信
- 实现不同host之间网络guest 互通
- 方便guest 迁移
- 支持网络数量扩大
- …

GRE tunnel 的不足

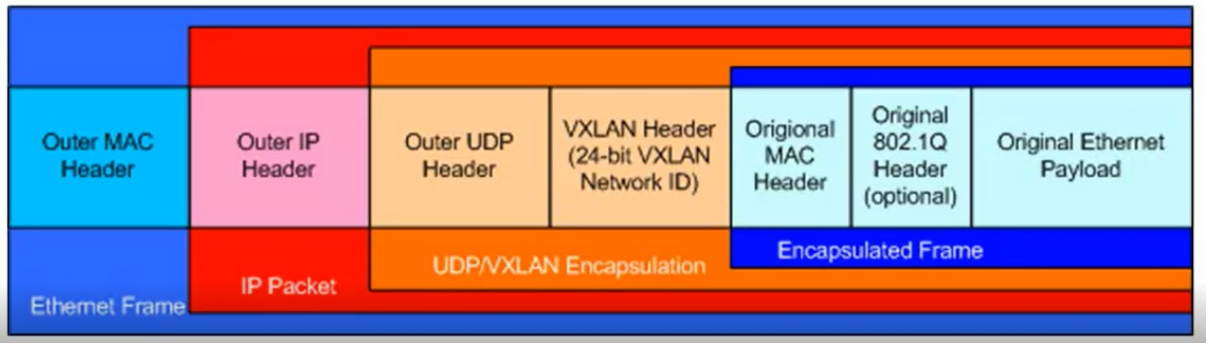
- 大规模部署问题
- 性能问题



随着物理主机的增加隧道会成n的n-1次方增长，网络隧道建的非常复杂，每个主机上面都会往所有的其他主机上建隧道。
它是在直接IP包上面跑IP包，当虚拟机接收实收相当于剥两层皮，而且涉及地址学习，所以有性能问题。

vxlan 数据包

- IP中封装MAC
- L3上包装L2
- L2 over UDP
- 1600万个VXLAN网



三层上面包二层包，使用udp来跑的，因为udp可以广播和单播。这里面放的是原始的二层网络的包。

Linux Interface Type

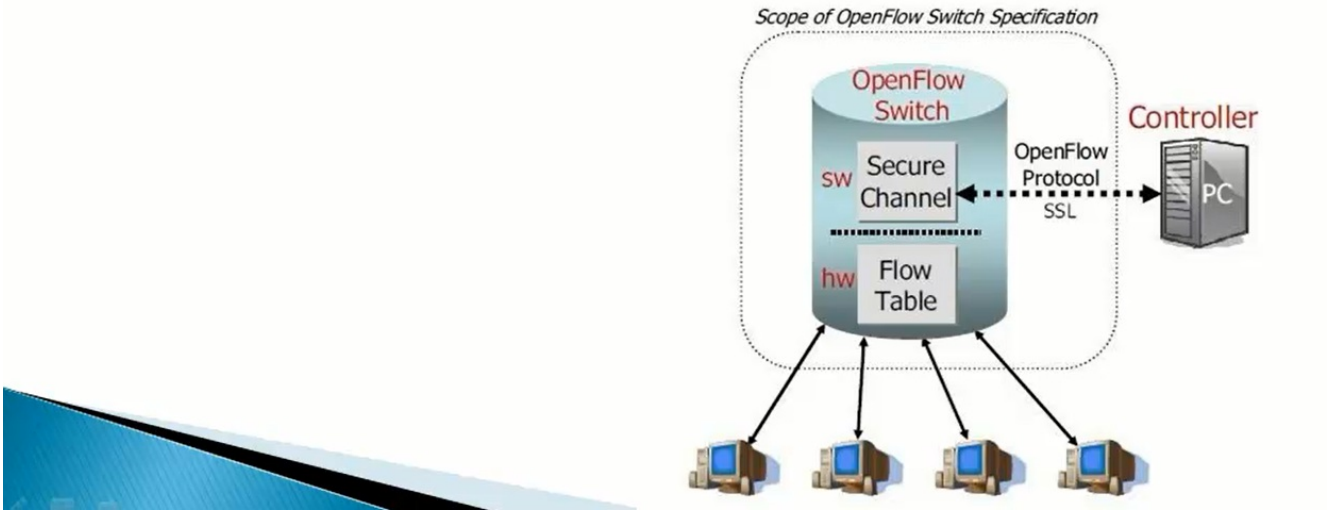
- TAP/TUN
- Bridge
- Physical
- Loopback

- Physical devices have a `/sys/class/net/eth0/device` symlink
- Bridges have a `/sys/class/net/br0/bridge` directory
- TUN and TAP devices have a `/sys/class/net/tap0/tun_flags` file
- Bridges and loopback interfaces have `00:00:00:00:00:00` in `/sys/class/net/lo/address`

TAP/TUN设备：是虚拟网卡的实现
Bridge：相当于物理中的交换机，连接所有的网卡和设备，他们之间收到的包都是一样的，每个网卡只抓到自己感兴趣的包。
Physical：物理网卡
Loopback：回环

Open vSwitch

- 网络隔离，vlan，gre，vxlan
- QoS配置
- 流量监控，Netflow，sFlow
- 数据包分析，Packet Mirror



专门做软交换机的项目，主要提供上述四种功能。因为是交换机所以提供网络隔离，同时支持三种隔离方式。同时可以提供Qos配置，对流量配置，也支持流量监控。