**ICP虚拟机网络不通排查思路**

**1、现象**

1. 虚拟机无法ping通，所有端口都不通。
2. 能ping通但是端口不通的，首先看下虚拟机使用的安全组，然后看下虚拟机里是否开了防火墙、端口是否监听，iptables是否放通等。其次考虑外部安全设备是否存在拦截。如果均没有，直接跳到第三步。

**2、虚拟机内部基础检查**

1. 查看虚拟机是否有平台分配的IP地址。没有IP时手动配置IP为平台分配的IP地址，可正常通信时再排查IP丢失原因。
2. MAC地址是否与平台的分配的一致。MAC地址必须与port的mac地址一致，不一致时无法正常通信。
3. 虚拟机网关是否配置，是否为平台子网中使用的网关。使用vxlan网络并且绑定了FIP的环境，网关必须配置为平台分配的网关，否则FIP无法通信。
4. windows虚拟机必须关闭防火墙；有安全软件的建议关闭后进行测试。
5. linux虚拟机查看防火墙配置是否有禁ping及drop对应请求的规则；iptables是否放通。

**3、虚拟机安全组检查**

vlan网络可以提供IP或者虚拟机ID；vxlan提供虚拟机ID

vlan网络提供了虚拟机IP地址时查询方式如下。

openstack server list --all --ip <IP地址>  
openstack port list --server <ServerID>  
openstack port show <PortID>  
openstack security group rule list --long <securityGroupID >

**4、链路检查**

**4.1 qbr桥检查**

* 找到虚拟机所在的计算节点和instance\_name

openstack server show <ID>  
​

* 登录到虚拟机所在计算节点，查找虚拟机使用的网桥

docker ps | grep libvirt.sh  
docker exec -it <CONTAINER ID> bash  
virsh domiflist <instance\_name>  
exit  
​

* 检查qbr桥信息是否连接正常

brctl show qbr<ID>  
​

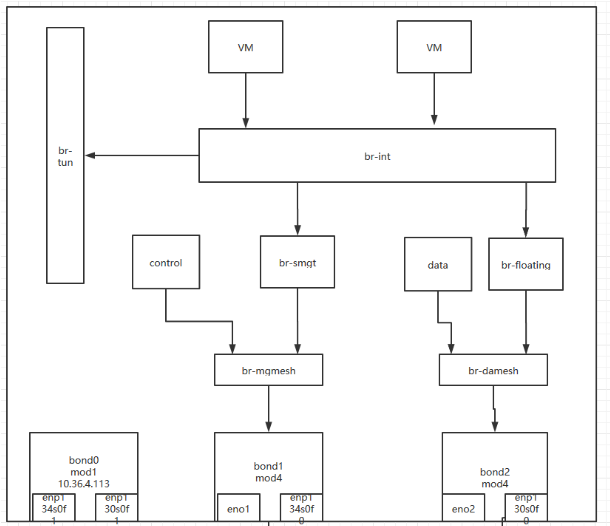
* 检查qvo桥信息是否连接正常

ovs-ofctl show br-int | grep qvo<ID>  
​

正常连接时qbr桥上会有两个口，一个是tap<ID>，一个是qvb<ID>。如果两个网口均能看到，则初步认为正常。 异常情况：缺少任意一个口，基本可以认为是异常情况，常见于ovs-agent异常，没有创建成功或添加成功端口。 解决方法：确认ovs-agent、ovs正常时，硬重启虚拟机；无法确认是否正常，执行热迁移或冷迁后到其他节点开机。

**4.2 ovs桥连接检查**

MCP环境版本或一些异常情况会导致ovs桥连接异常。正常情况下，桥的连接方式是这个样子：



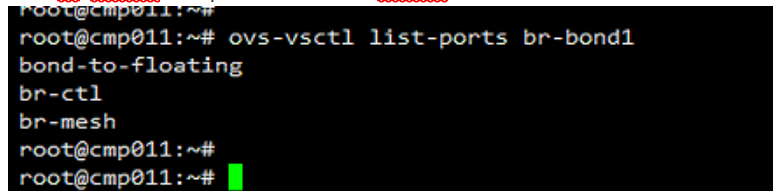
检查桥信息是否完善：

ovs-vsctl list-br  
​

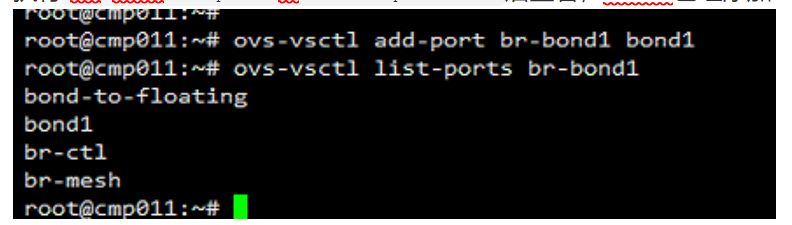
检查桥上的接口：

ovs-vsctl list-ports <br name>  
​

* 异常情况： 桥上缺少了对应的接口：常见的是br-mgmesh、br-damesh上没有bond接口，可以通过ovs-vsctl add-port <br name> <port name>进行添加。在计算节点上直接添加需要确认好，不然会有环路风险。 补充检查：在一些场景中执行ovs-vsctl list-ports <br name>显示可能是正常的，还需要检查下ovs-vsctl show中是否存在Error或Not Found等异常字样，如果存在异常，则需要在确认无误后执行手动添加操作。
* 生产环境实操(MCP环境)： 查看br-bond1上的port，发现缺少bond1



* 执行ovs-vsctl add-port <br name> <port name>后查看，bond1已经添加：



**5、网络抓包**

**5.1 发起测试请求**

能让一线配合测试时尽量由一线配合进行。如果一线不能配合，则使用如下方法：

**5.1.1 同网段发起测试**

同网段发起测试的主要在qdhcp ns中进行，如果vlan网络没有开启dhcp，则无法进行，必须由一线配合测试。 通过步骤3的openstack port show <PortID>方法找到虚拟机所属的网络ID。 登录到mgt04/05/06任一节点（网络节点）：

ip netns | grep <网络ID>  
ip netns e qdhcp-网络ID <执行所需的操作>  
​

通常可能执行的操作：

ping:  
ip netns e qdhcp-网络ID ping x.x.x.x  
​  
udp:  
ip netns e qdhcp-网络ID echo '123' >>/dev/udp/x.x.x.x/port  
eg:  
ip netns e qdhcp-网络ID echo '123' >>/dev/udp/192.168.1.5/666  
​  
telnet  
ip netns e qdhcp-网络ID telnet x.x.x.x port  
​

**5.1.2 不同网段发起测试**

登录到mgt04/05/06任一节点：

ip netns | grep snat  
选择任一snat的ns执行5.1.1中可能执行的操作  
​

**5.2 bond口抓包**

在bond口主要判断请求是否有到达计算节点，可以初步定为问题归属（平台网络/物理网络） 环境中绝大多数问题都在业务网卡，也就是bond2，这里的抓包网卡需要根据实际需求来判断，默认bond2。 抓所有报文：(量特别大，通常不建议使用这种方法。在不确定报文时，可以控制下抓包个数。在方法中增加-c 1000参数，1000为所需的抓包个数)

tcpdump -ne -i bond2  
​

抓取某个vlan的报文：

tcpdump -ne -i bond2 vlan <vlanID>  
​

抓取某个IP的报文：

tcpdump -ne -i bond2 host <IP地址>  
​

抓取某个端口的报文:

tcpdump -ne -i bond2 port <port>  
​

抓取某个协议的报文：

tcpdump -ne -i bond2 <协议>  
​

以上的都是基础方法，可以进行租户并添加部分处理逻辑，可以根据需求使用，eg:

tcpdump -ne -i bond2 vlan 200 and host 192.168.1.4 and port 666 and not src 192.168.1.5  
​

**5.3 linux桥抓包**

桥抓包主要在tap<ID>、qvb<ID>上进行。 如果在qvb<ID>上有请求，但是没有在tap<ID>上抓到，则可能是安全组导致；如果安全组配置允许通过，则可能是某段时间ovs-agent异常导致没有同步安全组配置，在确认平台各组件正常后可重启该节点上的ovs-agent进行恢复。 如果在tap<ID>上可以抓到请求，但是没有对应的响应报文，可以初步判断为虚拟机内部问题。

**6、问题排查样例**

**6.1 虚拟机无法与FIP通信**

**6.1.2 背景及现象**

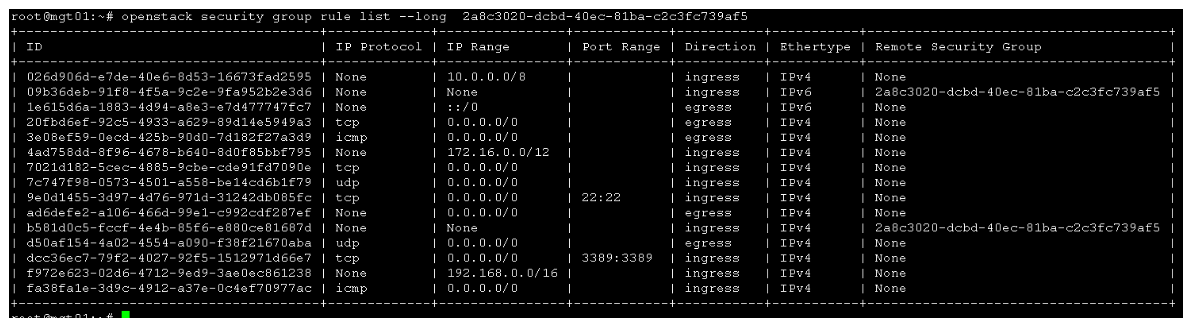
一线一批创建了五台vxlan的虚拟机，分别绑定了FIP，但是无法通过FIP进行通信，也无法ping通自己的FIP地址。

**6.1.3 基础检查**

虚拟机桥接部分正常



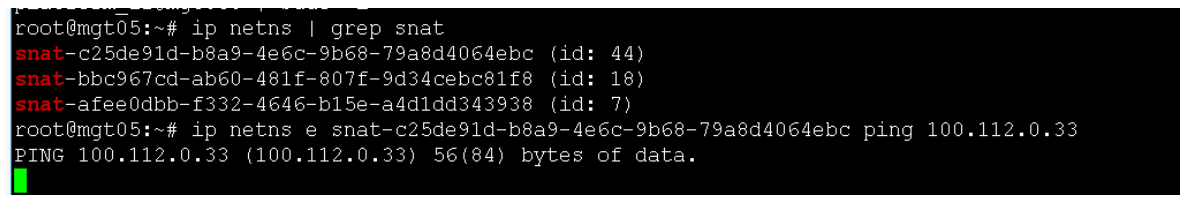
安全组放通了icmp协议



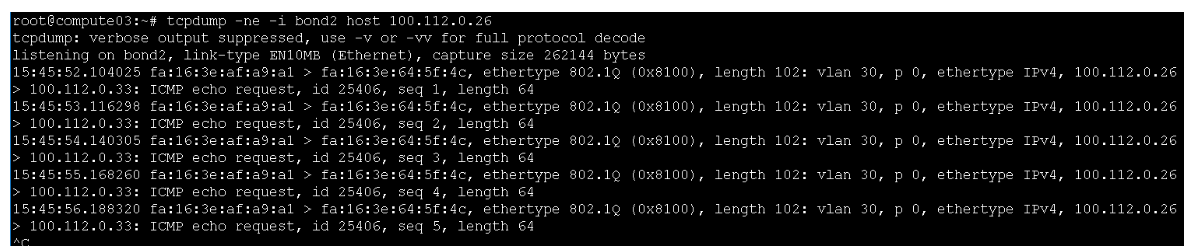
**6.1.4 同网段发包测试**

可以将FIP网络认为是一个普通的vlan网络。

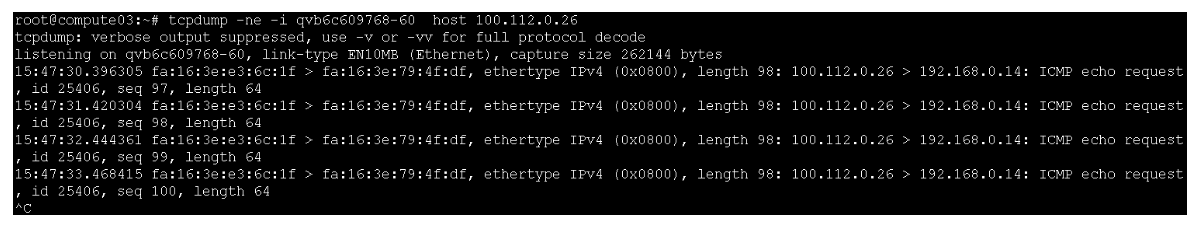
snat的namespace中发起到虚拟机FIP的ping测试。源IP为100.112.0.26，这个IP可以在snat的namespace中执行ip a获取，就是qg-<ID>设备上的IP。



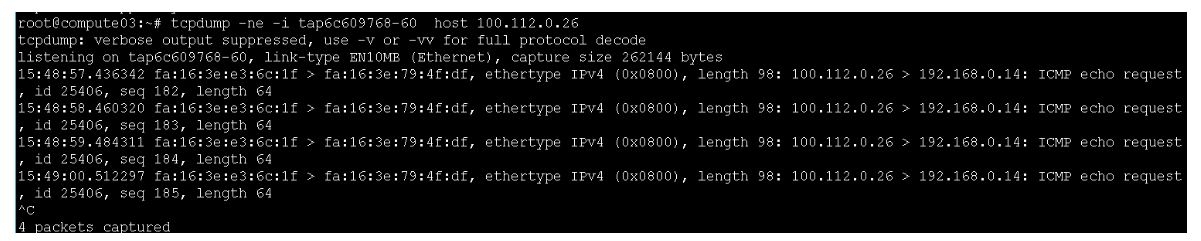
在虚拟机所在宿主机抓包，bond2口有ping请求，但是没有回应，初步排除物理网络问题。



在虚拟机使用的qvb-<ID>口抓包，可以看到ping请求：

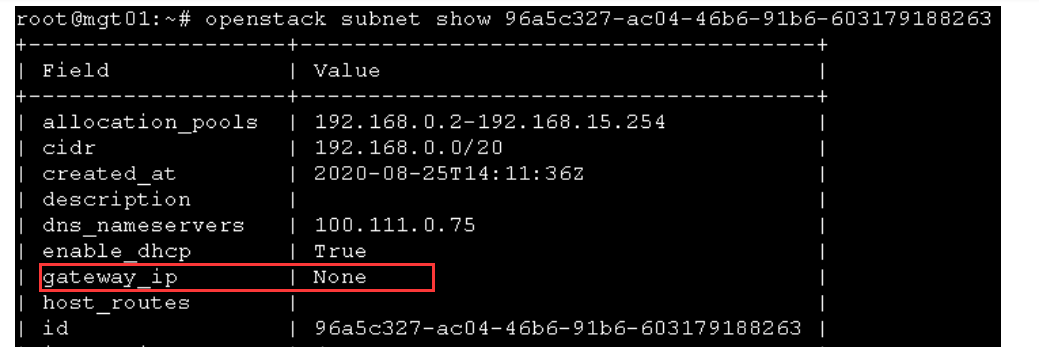


在虚拟机的tap-<ID>口抓到有ping请求，排除虚拟网络问题。



**6.1.5 扩展**

在排除平台虚拟网络后，可不再进行处理。后续工作非平台工作。 如果要登录到计算节点内排查，需要有明确的允许授权。 该问题是虚拟机内配置配置网关，继续排查发现subnet没有配置网关。



**6.2 某台虚拟机能ping通但是ssh时通时不通**

**6.2.1 背景及现象**

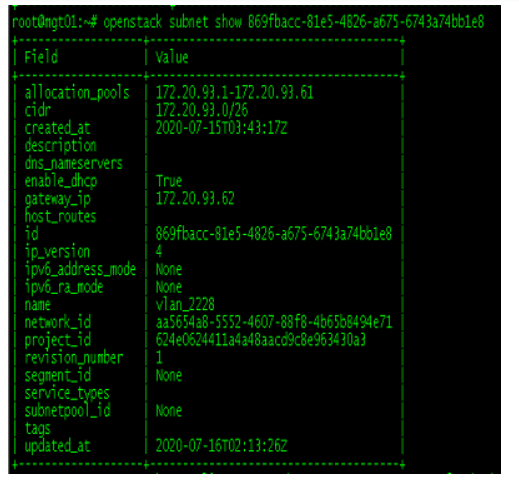
一线反馈某台虚拟机在使用一个vlan的某个port时会导致其他虚拟机无法和他通信，同一个虚拟机换成其他vlan的port就可以正常通信。

**6.2.3 基础检查**

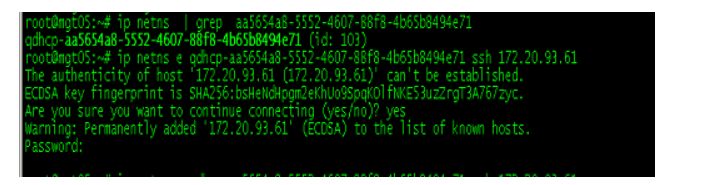
虚拟机网桥正常

**6.2.4 同网段发包测试**

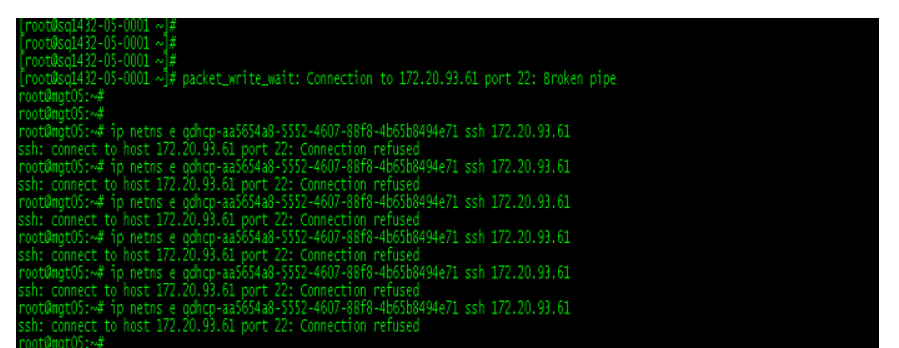
虚拟机子网开启了dhcp，可以在同网段发起测试



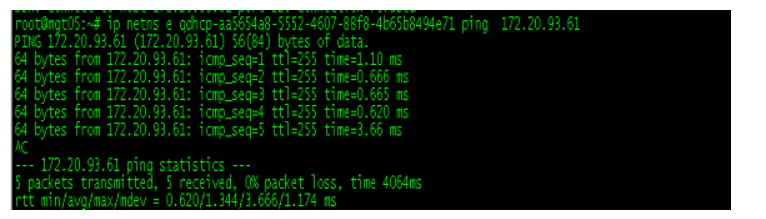
登录到mgt05，使用`ip netns | grep <network ID>`的方法获取到dhcp的namespace；在dhcp namespace中可以正常连接



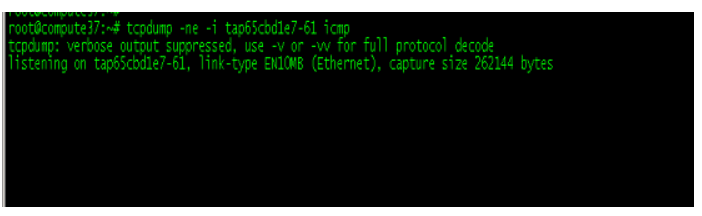
在连接一段时间后ssh断开，并且无法再ssh上去：



在ssh连不上时，ping仍然可以通信：

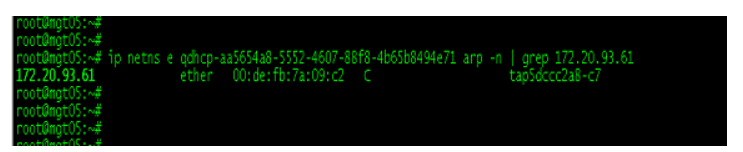


但是在计算节点上抓不到到虚拟机的ping包，由此初步判断IP地址冲突

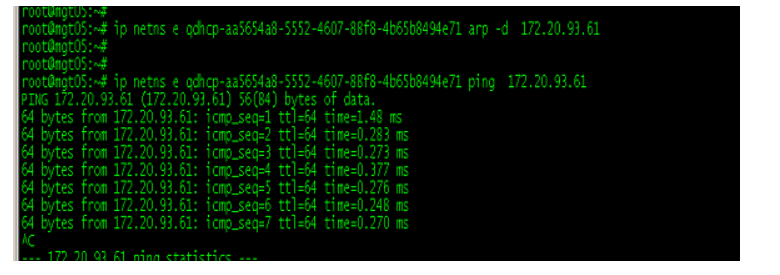


查看namespace中的arp，发现IP学习到的mac地址不是平台的port mac地址。

平台上的mac地址都是“fa:16”开头



在清理掉学习的错误mac地址后可以ping通，并且虚拟机上抓到了ping包，由此确认是IP地址冲突。



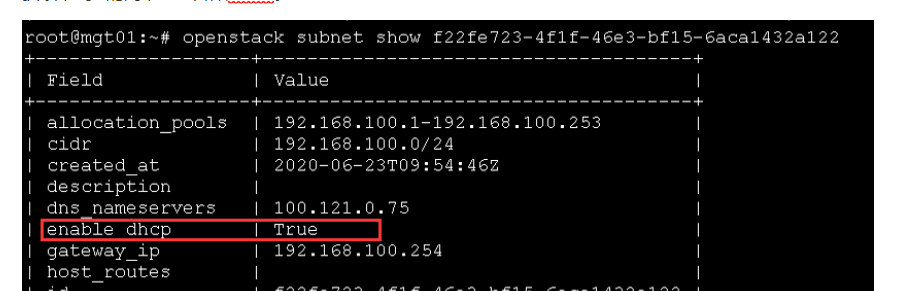
**6.3 虚拟机无法dhcp获取到IP地址**

**6.3.1 背景及现象**

一台虚拟机有两张网卡，有一个网卡通过dhcp获取不到IP地址。

**6.3.2 基础检查**

虚拟机网桥正常； 获取不到IP的网卡已经开始DHCP。

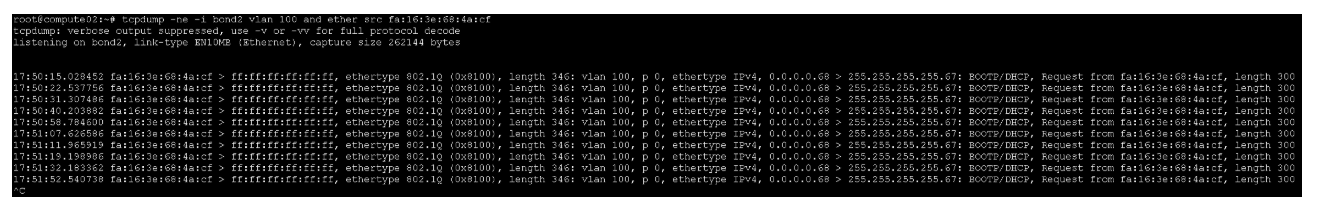


**6.3.3 抓包测试**

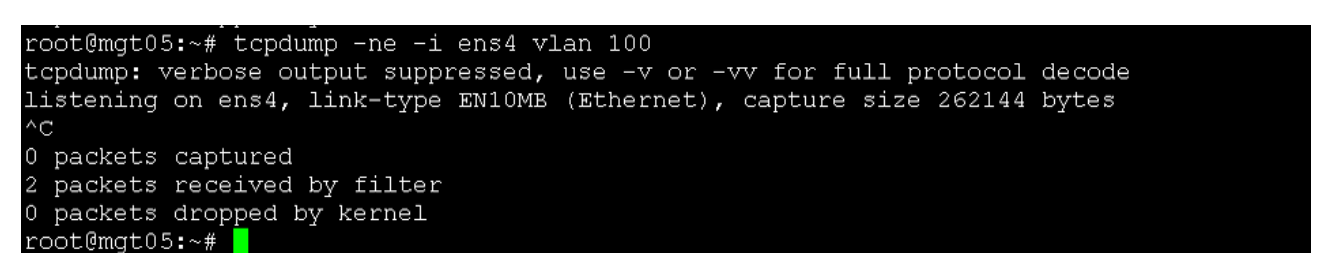
在虚拟机内执行dhclient,触发dhcpdiscover过程：



发现在bond2上有虚拟机发出的dhcp请求报文，但是没有应答。由此可知道虚拟机链路正常，基本可以排除虚拟机网络问题。



为排除其他问题影响，在mgt05节点没有收到广播过来的vlan 100的报文，基本可以确认物理网存在异常，需要物理网同事排查



**##问题**

* 1. window虚机网卡信息丢失，ip丢失

解决方法：虚机硬重启，重新挂载网卡。

Openstack server reboot <serverid> --hard

openstack server add port <server> <port>

* 1. vpc网络vxlan的虚机（重置系统操作后），dhclient eth0服务无法获取ip

定位思路：

虚机所在节点cmp157 ，查看端口状态为down（重置系统操作会导致虚机所在cmp节点变动，可能端口还留在之前的节点）

宿主机 ovs-ofctl dump-flows br-int 未见卡死，neutron日志未见 “OVS is dead”。

虚机内部dhcpclient eth0 ，在cmp157节点抓包，未抓到包 （tcpdump -i bond2 ether src fa:16:3e:14:d5:44）（vxlan网络的虚机无法在cmp节点抓包）

解决方法：对虚机进行热迁移

* 1. 同网段虚机互ping不通，查看虚机ip丢失

systemctl reatart neutron-openvswitch-agent.service

ovs-ofctl dump-flows br-int 卡死  
原因：br-int上table61缺少一条热submit至62的流表，重启ovs-agent后恢复。该问题已经在cmp098上出现过一次，具体原因不清楚，请L3排查  
日志中均出现过“OVS is dead”，但是理论上并不会丢失该部分流表，请一并看下ovs异常

* 1. staging租户侧虚机网卡没有获取到IP

在虚拟机内执行dhclient eth1,触发dhcpdiscover过程。

在所在compute030节点抓包(tcpdump -ne -i ten1 vlan 3012)，发出的dhcp请求报文，但是没有应答。

在gtw节点没有收到广播过来的vlan 3012的报文，基本可以确认物理网存在异常，需要物理网同事排查。

应该是compute030节点没有接入vlan 3012

vlan128的问题，认为交换机问题：从虚机内部ping网关，arp包从compute节点的bond2出去了，但是没有返回，没有学到mac。

* 1. ntp同步问题

先检查ntpdate工具能否安装成功

* zww区发包测试：

echo '123' >>/dev/udp/xx.xx.xx.xx/123

或者

ntpdate -d xx.xx.xx.xx

* hlw区抓包:

tcpdump -ne -i ens4.3108 port 123

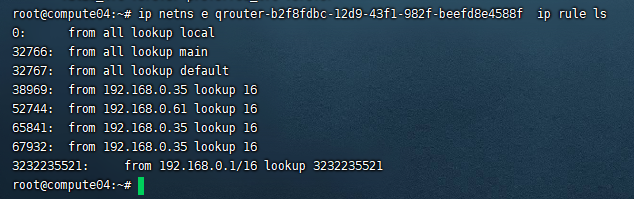
* 1. 同一集群内的虚机vpc网络，有的可以通floating网关，有的不行

https://blog.csdn.net/u012758088/article/details/76255543

虚机所在compute节点，绑定float的才会有rule，未绑定的不应该存在，需要删除

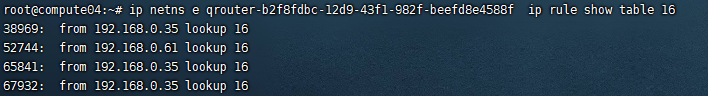
ip netns e qrouter-xxx ip a

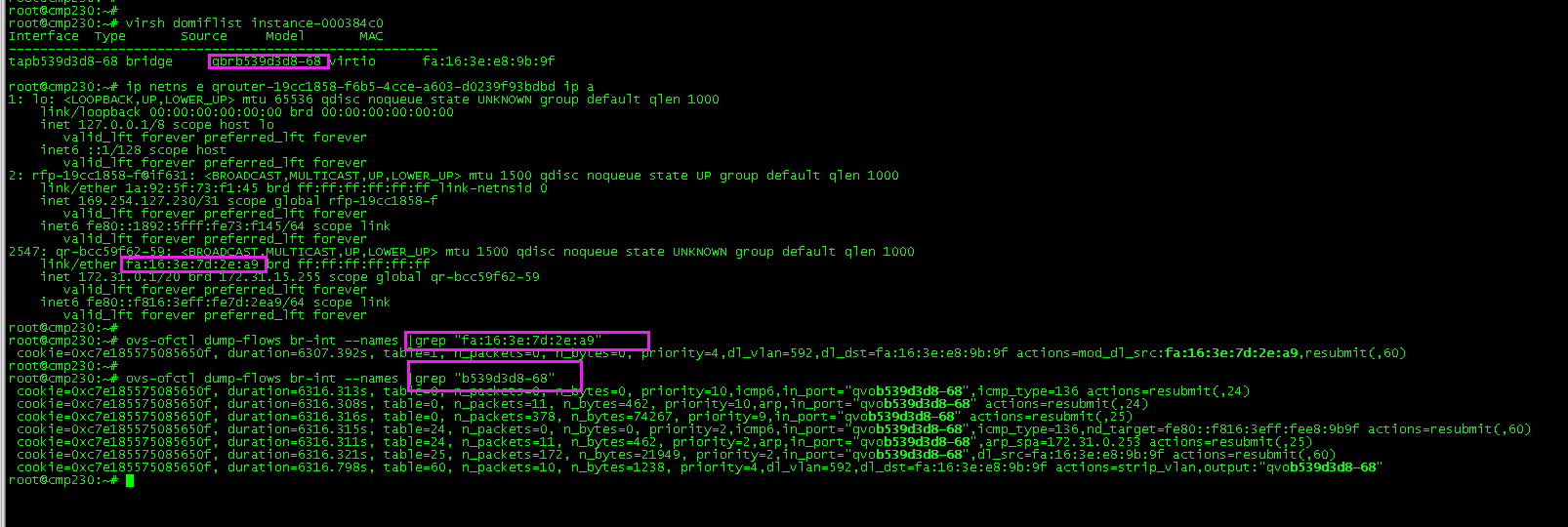
ip netns e qrouter-xxx ip rule show



ip netns e qrouter-xxx ip rule show table 16

ip netns e qrouter-xxx ip rule show table main





ip netns e qrouter-b2f8fdbc-12d9-43f1-982f-beefd8e4588f bash

ip netns e qrouter-b6a7dce8-4a97-481c-ae05-205ceb83da31 ip rule del prio 42570

添加路由：





* 1. 网络连通，但数据库和应用服务器无法ssh连接不了

解决方法：虚机所在宿主机cmp044 上数据网卡bond1（控制网） 的mtu是1500，正常应该是9000，已调整。

* 1. 云主机扩容后，内网IP无法正常访问

解决方法： 查看虚机的port 状态是down，手动启动

openstack port set --enable 83dfd0e8-954e-425f-ad1e-88f0d14264df

* 1. 更换公网EIP后，无法连接数据库；机器内无内网ip；远程登陆失败，ping不通，22端口已经开了；无法ping通网关；客户反馈这个EIP，之前能正常ping通过，突然无法ping通了

解决方法：热迁移操作

* 1. 信创云平台因业务不通重启租户虚拟机，目前重启进程一直卡在重启中，无法重启完成

定位方法：查看虚机所在节点stor-public网和其他节点以及网关的连通性

解决方法：重启stor-public的网卡后，网络连通后解决。

* 1. 华北三云主机扩容后无法ping通过网关

原因是cmp059节点qr-设备vlan号没有更新成功，已手动改好，产生原因还需要排查

* 1. 安全组开通端口38069、38443，但是实际端口不通

解决方法：客户自己应用问题

* 1. 华北ECS登录失败，找不到网卡，重启进入救援模式

网卡down 掉了；进入系统，查看有/dev/vdd /data的挂载点，而盘中没有/vdd，500G的数据盘是/dev/vdc，麻烦注释/ect/fstab中/dev/vdd /data该行，再重启虚拟机

* 1. 客户反馈远程的登陆不上，客户还反馈6389,8090端口不通。

执行pam\_tally2 --user root --reset，清零root用户失败计数；

* 1. 172.31.0.55的虚机无法ssh到同一VPC下172.31.0.39的虚机，反向则可以，与同一VPC下的其他虚机ssh连接也没有问题，

解决方法：所在宿主机cmp501计算节点的数据网卡ten3的mtu设置的有问题，已调整为9000

* 1. BSS后台staging环境登录不上

网络上禁用了从办公网直连10.220段