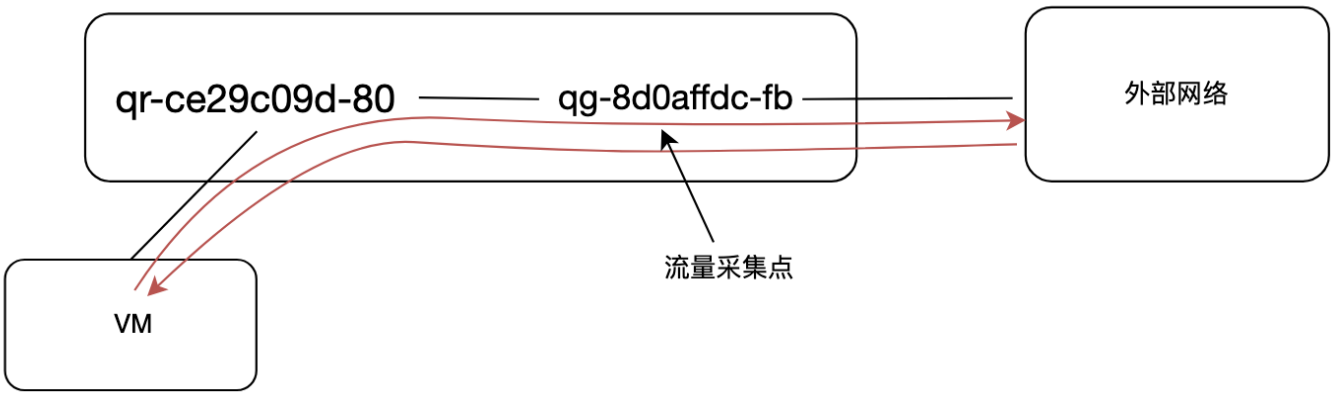
iptables流量统计

iptables准确性验证

虚拟机进出外网流量路径如下



iptables采集规则如下， 10.0.1.122是虚拟机内网IP，38.175.42.245是浮动IP

```
iptables -t mangle -A PREROUTING -d 38.175.42.245 -i qg-586a8fbb-0f -m comment --comment traffic_in_38.175.42.245  -j ACCEPT
iptables -t mangle -A POSTROUTING -s 10.0.1.122 -o qg-586a8fbb-0f -m comment --comment traffic_out_38.175.42.245 -j ACCEPT
```

主要在mangle表中通过在PREROUTING，POSTROUTING两个链中插入规则，去统计进出流量。

测试结果

虚拟机流量查看，通过ifconfig观察网卡rx/tx计数器

采集点流量查看，通过iptables -L -nvx 查看计数器

方向	数据量	iperf速率(Mb/s)	虚拟机	采集点	误差	误差率	理论最低值(payload)	理论最高值(payload + headers)
进（公网进虚拟机的流量）	10G	100	10755383341	10752313814	3069527	0.02%	10737418240	11031594080
		50	10769500882	10767186842	2314040	0.02%		
		20	10754641725	10750649996	3991729	0.03%		
		10	10768927837	10766148157	2779680	0.02%		
	1000G	1000	1076213272247	10752517325 01	961539746	0.08%	10737418240 00	11031594080 00
		500	1075754311683	10755901576 39	164154044	0.02%		
出（虚拟机出公网的流量）	10G	100	10753879826	10750507363	3372463	0.03%	10737418240	11031594080
		50	10753979367	10750554539	3424828	0.03%		
		20	10753983924	10750577525	3406399	0.03%		
		10	10754126919	10750664799	3462120	0.03%		
	1000G	1000	1075367474601	10750305560 20	336918581	0.03%	10737418240 00	11031594080 00
		500						

iperf测试命令 iperf3 -c 38.175.42.245 -n 10G -f 50M

准确性总结

根据以上表格的信息，可以看到iptables跟虚拟机网卡的收发计数器存在误差，误差原因如下：

1、iptables不会计算以太网头部信息，所以每个报文少14个字节。

2、网卡TSO/GRO/GSO等特性，导致内核合并报文，如果一次性合并40个报文，那么iptables会少计算1600字节，1600 = 40 * (20(ip) + 20(tcp))。

表格中的数据也确实说明iptables少计算了数据量。

由于实际环境中，每个环节报文的封装、MTU、网卡特性等信息不同，所以理论最低值应该抛开数据包各种头部信息，以实际payload数据量为准，只要iptables统计的数据不低于这个理论最低值，不高于加上头部信息的理论最高值，那结果就是准确的。

另外一个需要注意的点是，丢包引起的准确性降低，这种情况下，可能会导致iptables数据便高也可能便低，取决于实际场景。

openstack流量统计场景

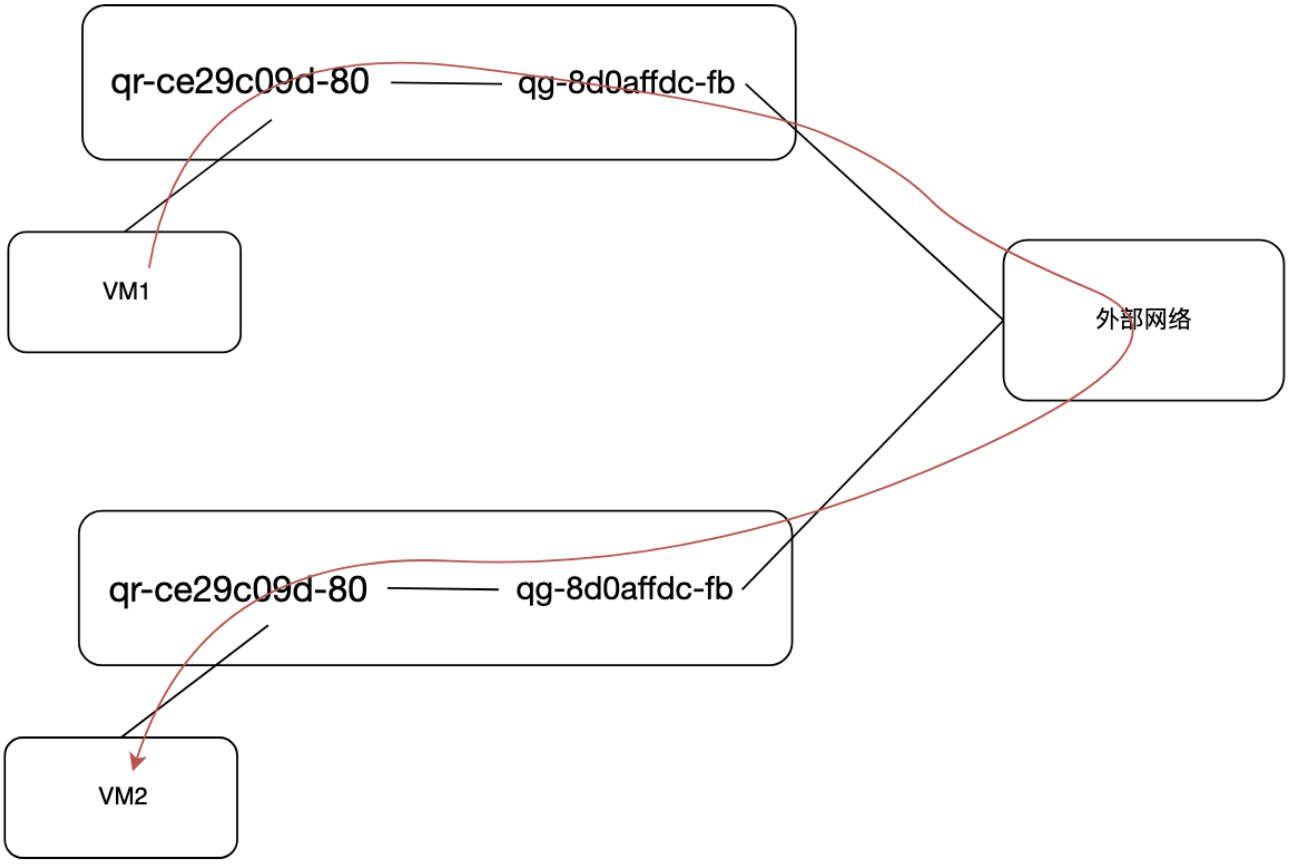
场景1，虚拟机进出外网

测试没问题。

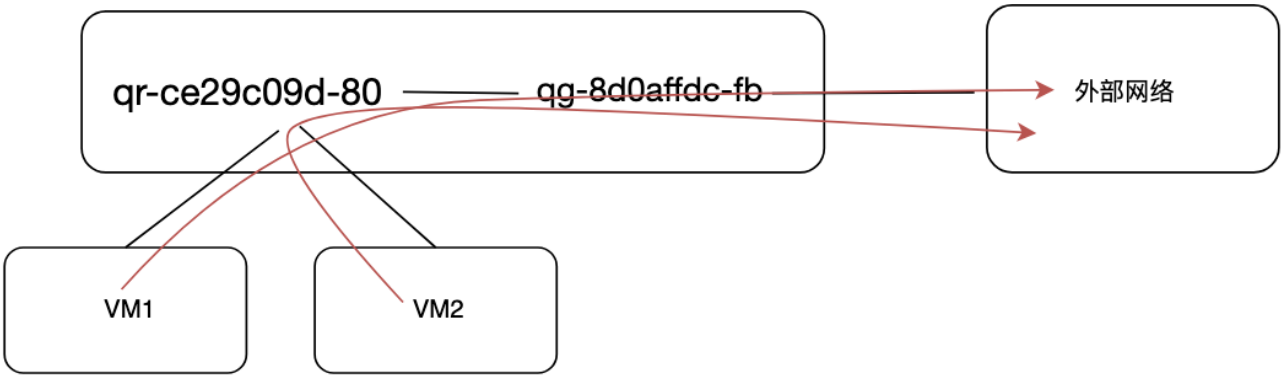
场景2，虚拟机通过浮动IP访问云内资源，同一个vrouter

使用"场景1"中的采集规则无法100%满足，原因是同一个vrouter下的虚拟机通过浮动IP通信时，流量不会经过qg-接口，如下

场景3，虚拟机通过浮动IP访问云内资源， 不同vrouter



场景4，虚拟机没浮动IP的情况



场景5，虚拟机没有浮动ip的情况，访问云内资源，同router

同场景2

场景6，虚拟机没有浮动ip的情况，访问云内资源，不同router

同场景3

场景7，公网IP限速的时候，数据是否准确

公网IP在qos限速的情况下，iptables可以获取到收发数据。

场景8，端口映射

国际站没有这个功能，暂时不考虑

iptables采集规则

```
ipset create subnet1 hash:ip
ipset add subnet1 X.X.X.X  # X.X.X.X为没有浮动ip的虚拟机内网ip地址
路由器浮动ip流量采集
iptables -t mangle -A PREROUTING -m set --match-set subnet1 src -m set ! --match-set subnet1 dst -m comment --comment traffic_snat_10.11.4.217_out -j traffic_chain
iptables -t mangle -A PREROUTING -m set -d 10.11.4.217/32 ! --match-set subnet1 src -m comment --comment traffic_snat_10.11.4.217_in -j traffic_chain
虚拟机浮动ip流量采集
iptables -t mangle -A PREROUTING -s 10.0.1.134/32 -m set ! --match-set subnet1 dst -m comment --comment traffic_10.11.4.164_vm_out -j traffic_chain
iptables -t mangle -A PREROUTING -d 10.11.4.164/32 -m comment --comment traffic_10.11.4.164_vm_in -j traffic_chain
```

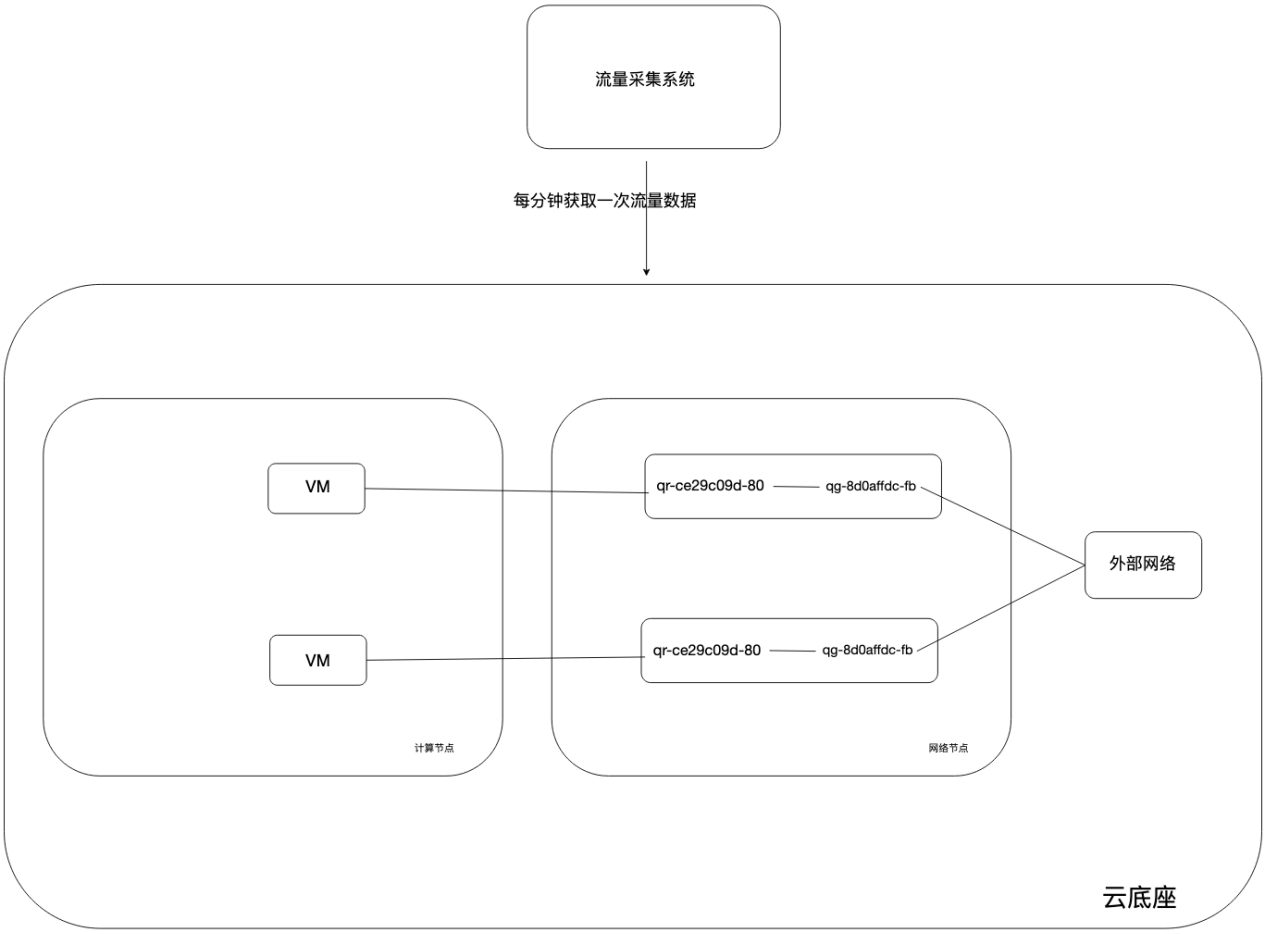
以上规则能满足 场景1 - 场景7 下的浮动IP流量采集。

路由器，子网，端口，浮动IP等有变动时，需要调整iptables相关规则。

采集间隔

1分钟

整体采集架构



iptables工具

rocky8.6默认使用的是nftables，openstack容器里面使用的是iptables，这个导致在宿主机上使用 iptables 命令无法看到规则。

问题

场景3，qos限速会失效（这个不属于流量采集问题）

一些流量不应该被计费，如169.254.0.0/16相关流量