

知 MSR V7系列路由器做到来回路径一致的经验案例

NAT 马文斌 2015-11-09 发表

一、 组网：

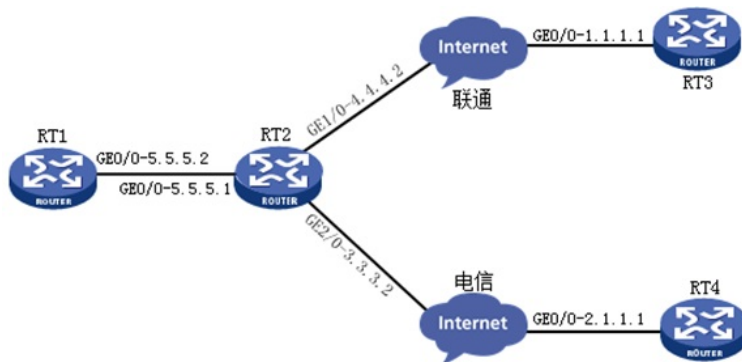


图1

如图所示，RT2上为双出口，一个联通，一个电信，在内网要开放一台telnet服务器，公网中的RT3、RT4设备分别位于联通和电信的网络中，客户要求，当RT3通过RT2的GE1/0接口访问RT1的时候，从RT1发给RT3的报文也必须通过RT2的GE1/0接口回去，同理，RT4通过RT2的GE2/0接口访问内部RT1的时候，RT1回复给RT4的报文，也必须通过RT2的GE2/0接口回去。

二、 问题描述：

涉及到2个问题：

问题一：

MSR V7设备不像SR66设备一样，可以在策略路由中配置if-match reverse-input-interface命令用来设置反向入接口匹配规则，这样就可以做到来回路径一致，但是MSR设备上没有这个命令，应该怎么做呢？

问题二：

按照下边方法配置完成后，发现已经形成了nat session，但是和内网服务器不通。

三、 过程分析：

【问题一】

先来看看这个问题是如何引入的：

如图所示，RT2作为NAT路由器有2个出口，分别为联通和电信（实际中也有可能是同一个运行商的2条不同线路），公网上2个网段，RT3和RT4，在RT2上两个公网口同时要映射到内网口同一个RT1的服务，也是就要在RT2两个公网口上配置nat server。

RT3通过访问RT2的联通地址，RT2会通过nat server映射到RT1上，此时RT2会将目的地址转换为RT1的地址和端口，送给RT1，但是RT2不会转换源地址，然后RT1回复报文，目的地址是RT3，源地址是RT1，送给RT2的内网口，此时RT2要查找路由，决定将此报文通过哪个接口送出去，通常这种情况客户在RT2上配置的都是等价路由，而RT2收到的目的地址又是公网地址，所以这种情况下，RT2就有可能将报文通过电信接口送出去，如果通过电信接口送出去的话，就不会匹配原来的nat session，导致访问不通。

知道了此问题的原因，我们可以通过nat inbound技术来解决这个问题。

nat inbound原理为：在报文进入接口的时候匹配nat inbound，改变报文源地址，知晓了nat inbound的作用之后，再来细细分析这个问题的根本原因：

这个问题的根本原因为，RT1回复的目的地址是公网地址，由于RT2没有明细路由，所以导致报文匹配默认路由出去，而默认路由通常是等价的，设备会通过哈希的方法决定报文的去处，所以无法保证报文来回路径一致。

那么问题逐渐清晰了，想要解决这个问题，只需要在RT2接收公网报文的时候，在源地址上做文章就可以：

配置的思路为：要在RT2的出口上同时配置nat inbound和nat server即可，nat inbound转换报文源地址，nat server转换报文目的地址。

那nat inbound如何配置呢？

推荐配置为公网接口的下一跳地址，比如，RT2上的联通地址为4.4.4.2，下一跳为4.4.4.1，那么nat inbound的地址-group中就配置4.4.4.1，为何这样配置？

上边已经提到，RT2收到内网口的报文，会查找路由来决定送给那个接口，而相对于公网网关而言，都是直连路由，就不必再通过手工增加路由了，从而就可以保证回去的报文，肯定会从进来的接口回去。

如果配置的不是公网下一跳地址，会造成什么问题呢？

【问题二】

这就是问题二，会造成RT2可以形成nat session，源地址和目的地址都会被转换，但是业务不会通，原因就是路由问题，上述比较清楚了，这里就不再赘述。

那这里又引出一个问题，如果配置的nat inbound的地址-group中的地址和出口为同一个网段，但是不是网关地址，不配置路由，是否可以？

答案是不可以，我们都知道，路由器在查找完路由表之后，会查找ARP表封装下一跳的mac地址，但是ARP中不会有这个表现的，举个例子说明：

如图中所示：联通的地址为4.4.4.2，如果nat inbound的地址-group中配置的地址为4.4.4.3，而非网关4.4.4.1，那么当RT1回复目的地址为4.4.4.3到RT2后，RT2确实可以找到直连路由4.4.4.0/24的直连路由，但是当在ARP表中查找4.4.4.3的时候，是没有的，此时RT2会发送ARP报文询问4.4.4.3的MAC地址是谁，4.4.4.3在网络中是不存在的，所以肯定没有人回应，找不到ARP，也就无法进行下一步转发，路由器会丢弃此报文，所以会造成业务不通。

如下：

```
*Oct 26 07:55:31:844 2015 H3C ARP/7/ARP_SEND: Sent an ARP message, operation: 1, sender MAC: 52c8-ae67-0306, sender IP: 4.4.4.2, target MAC: 0000-0000-0000, target IP: 4.4.4.3
```

所以路由是必须增加的，增加路由后，大家都知道，RT2查找路由找到下一跳之后，自然会查找下一跳的ARP表项，下一跳是直连路由，自然会有ARP。

四、 解决方法：

问题一的解决方法已然很明了，只要在内网口同时配置nat inbound和nat server原理上即可实现客户需求。

问题二实际上涉及两种情况，分别分析：

情况一：nat inbound的地址-group中配置的地址和出接口不在同一个网段中。

解决方法只能是配置静态路由，将address-group中的地址段指向公网网关

情况二：nat inbound的地址-group中配置的地址和出接口在同一个网段中。

解决方法有两种：

- 一、配置静态路由，将address-group中的地址指向公网网关
- 二、配置静态ARP，将address-group中的IP地址和公网网关的mac地址关联起来。

下面就几种情况做个测试：

1、address-group中配置和接口地址不在同一个网段：

【RT1配置】

```
#
telnet server enable //开启telnet服务
#
interface GigabitEthernet0/0
port link-mode route
combo enable copper
ip address 5.5.5.2 255.255.255.0
#
ip route-static 0.0.0.0 0 5.5.5.1
```

【RT2配置】

```
sysname H3C
#
```

```

interface GigabitEthernet0/0 //连接内网RT1的接口和地址
port link-mode route
combo enable copper
ip address 5.5.5.1 255.255.255.0
#
interface GigabitEthernet0/1 //连接联通的地址和接口
port link-mode route
combo enable copper
ip address 4.4.4.2 255.255.255.0
//配置nat inbound使用acl做到精细控制，并关联address-group组
nat inbound 3000 address-group 1
//配置nat server，将当前接口地址和23端口映射到内网RT1的地址和23端口
nat server protocol tcp global current-interface 23 inside 5.5.5.2 23
#
interface GigabitEthernet0/2 //连接电信的地址和接口
port link-mode route
combo enable copper
ip address 3.3.3.2 255.255.255.0
//配置nat inbound使用acl做到精细控制，并关联address-group组
nat inbound 3000 address-group 2
//配置nat server，将当前接口地址和23端口映射到内网RT1的地址和23端口
nat server protocol tcp global current-interface 23 inside 5.5.5.2 23
#
//配置到address-group地址的路由，下一跳必须要配置正确
ip route-static 200.1.1.1 32 3.3.3.1
ip route-static 100.1.1.1 32 4.4.4.1
#
acl advanced 3000 //关联nat inbound的acl，做到精细化控制
rule 0 permit tcp destination-port eq telnet
#
nat address-group 1 //关联联通接口的address-group
address 100.1.1.1 100.1.1.1 //地址和接口不在同一网段
#
nat address-group 2 //关联电信接口的address-group
address 200.1.1.1 200.1.1.1 //地址和接口不在同一网段
【验证】
telnet 4.4.4.2
Trying 4.4.4.2 ...
Press CTRL+K to abort
Connected to 4.4.4.2 ...

```

```

*****
* Copyright (c) 2004-2014 Hangzhou H3C Tech. Co., Ltd. All rights reserved. *
* Without the owner's prior written consent,                               *
* no decompiling or reverse-engineering shall be allowed.                 *
*****

```

```
login: 123 //登陆成功

Password:

RT2上查看nat session:

display nat session verbose

Slot 0:

Initiator: //入方向为1.1.1.1访问4.4.4.2的23端口

Source    IP/port: 1.1.1.1/17222

Destination IP/port: 4.4.4.2/23

DS-Lite tunnel peer: -

VPN instance/VLAN ID/VLL ID: -/-/

Protocol: TCP(6)

Inbound interface: GigabitEthernet0/1

Responder: //出方向为5.5.5.2的23端口回复100.1.1.1

Source    IP/port: 5.5.5.2/23

Destination IP/port: 100.1.1.1/1029

DS-Lite tunnel peer: -

VPN instance/VLAN ID/VLL ID: -/-/

Protocol: TCP(6)

Inbound interface: GigabitEthernet0/0

State: TCP_ESTABLISHED

Application: TELNET

Start time: 2015-10-26 08:46:55  TTL: 3561s

Initiator->Responder:      0 packets      0 bytes

Responder->Initiator:      0 packets      0 bytes

Total sessions found: 1
```

2、address-group中配置和接口地址在同一个网段，但是和网关地址不一致：

2.1 使用路由方法配置：

RT2修改配置如下：

```
//配置到address-group地址的路由，下一跳必须要配置正确

ip route-static 3.3.3.3 32 3.3.3.1

ip route-static 4.4.4.3 32 4.4.4.1

#

acl advanced 3000 //关联nat inbound的acl，做到精细化控制

rule 0 permit tcp destination-port eq telnet

#

nat address-group 1 //关联联通接口的address-group

address 4.4.4.3 4.4.4.3

#

nat address-group 2 //关联电信接口的address-group

address 3.3.3.3 3.3.3.3

其他配置不变。

【验证】

telnet 4.4.4.2
```

Trying 4.4.4.2 ...

Press CTRL+K to abort

Connected to 4.4.4.2 ...

* Copyright (c) 2004-2014 Hangzhou H3C Tech. Co., Ltd. All rights reserved. *

* Without the owner's prior written consent, *

* no decompiling or reverse-engineering shall be allowed. *

login: 123 //登陆成功

Password:

RT2上的nat session信息:

[RT2]display nat session verbose

Slot 0:

Initiator: //入方向为1.1.1.1访问4.4.4.2的23端口

Source IP/port: 1.1.1.1/17223

Destination IP/port: 4.4.4.2/23

DS-Lite tunnel peer: -

VPN instance/VLAN ID/VLL ID: -/-/-

Protocol: TCP(6)

Inbound interface: GigabitEthernet0/1

Responder: //回复方向为5.5.5.2的23端口回复4.4.4.3

Source IP/port: 5.5.5.2/23

Destination IP/port: 4.4.4.3/1025

DS-Lite tunnel peer: -

VPN instance/VLAN ID/VLL ID: -/-/-

Protocol: TCP(6)

Inbound interface: GigabitEthernet0/0

State: TCP_ESTABLISHED

Application: TELNET

Start time: 2015-10-26 08:54:23 TTL: 3544s

Initiator->Responder: 0 packets 0 bytes

Responder->Initiator: 0 packets 0 bytes

Total sessions found: 1

查看IP路由表，可以看到到达4.4.4.3的精确路由

[RT2]display ip routing-table 4.4.4.3

Summary Count : 4

Destination/Mask	Proto	Pre	Cost	NextHop	Interface
0.0.0.0/0	Static	60	0	3.3.3.1	GE0/2
			4.4.4.1	GE0/1	
4.4.4.0/24	Direct	0	0	4.4.4.2	GE0/1
4.4.4.3/32	Static	60	0	4.4.4.1	GE0/1

查看arp表项，没有4.4.4.3的ARP

[RT2]display arp

Type: S-Static D-Dynamic O-Openflow R-Rule M-Multiport I-Invalid

IP address	MAC address	VLAN	Interface	Aging Type
------------	-------------	------	-----------	------------

3.3.3.1	52c8-a9a6-0205	N/A	GE0/2	11 D
---------	----------------	-----	-------	------

4.4.4.1	52c8-a203-0105	N/A	GE0/1	12 D
---------	----------------	-----	-------	------

5.5.5.2	52c9-341b-0505	N/A	GE0/0	13 D
---------	----------------	-----	-------	------

2.2 使用ARP静态绑定方法

修改RT2配置如下，其他配置不动：

删除静态路由

[RT2]undo ip route-static 4.4.4.3 32 4.4.4.1

[RT2]undo ip route-static 3.3.3.3 32 3.3.3.1

查看arp表项，找到网关IP地址对应的mac地址

[RT2]display arp

Type: S-Static D-Dynamic O-Openflow R-Rule M-Multiport I-Invalid

IP address	MAC address	VLAN	Interface	Aging Type
------------	-------------	------	-----------	------------

3.3.3.1	52c8-a9a6-0205	N/A	GE0/2	11 D
---------	----------------	-----	-------	------

4.4.4.1	52c8-a203-0105	N/A	GE0/1	12 D
---------	----------------	-----	-------	------

5.5.5.2	52c9-341b-0505	N/A	GE0/0	13 D
---------	----------------	-----	-------	------

将address-group中的地址和网关的MAC静态绑定起来

[RT2]arp static 4.4.4.3 52c8-a203-0105

[RT2]arp static 3.3.3.3 52c8-a9a6-020

【验证】

telnet 4.4.4.2

Trying 4.4.4.2 ...

Press CTRL+K to abort

Connected to 4.4.4.2 ...

* Copyright (c) 2004-2014 Hangzhou H3C Tech. Co., Ltd. All rights reserved. *

* Without the owner's prior written consent, *

* no decompiling or reverse-engineering shall be allowed. *

login: 123 //登陆成功

Password:

RT2上的nat session信息：

[RT2]dis nat session verbose

Slot 0:

Initiator:

Source IP/port: 1.1.1.1/17225

Destination IP/port: 4.4.4.2/23

DS-Lite tunnel peer: -

VPN instance/VLAN ID/VLL ID: -/-/-

Protocol: TCP(6)

Inbound interface: GigabitEthernet0/1

Responder:

Source IP/port: 5.5.5.2/23

Destination IP/port: 4.4.4.3/1027

DS-Lite tunnel peer: -

VPN instance/VLAN ID/VLL ID: -/-/-

Protocol: TCP(6)

Inbound interface: GigabitEthernet0/0

State: TCP_ESTABLISHED

Application: TELNET

Start time: 2015-10-26 09:13:47 TTL: 3566s

Initiator->Responder: 0 packets 0 bytes

Responder->Initiator: 0 packets 0 bytes

Total sessions found: 1

[RT2]display ip routing-static 4.4.4.3 //没有4.4.4.3的精细路由，只有直连路由

Summary Count : 3

Destination/Mask	Proto	Pre	Cost	NextHop	Interface
0.0.0.0/0	Static	60	0	3.3.3.1	GE0/2
			4.4.4.1	GE0/1	
4.4.4.0/24	Direct	0	0	4.4.4.2	GE0/1

[RT2]display arp //可以看到静态绑定的arp信息

Type: S-Static D-Dynamic O-Openflow R-Rule M-Multiport I-Invalid

IP address	MAC address	VLAN	Interface	Aging	Type
4.4.4.3	52c8-a203-0105	N/A	GE0/1	N/A	S
3.3.3.3	52c8-a9a6-0205	N/A	N/A	N/A	S
3.3.3.1	52c8-a9a6-0205	N/A	GE0/2	14	D
4.4.4.1	52c8-a203-0105	N/A	GE0/1	15	D
5.5.5.2	52c9-341b-0505	N/A	GE0/0	16	D