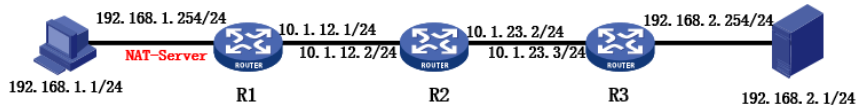
某局点MSR G2路由器Tracert显示异常经验案例

吕甲南　2017-07-12 发表

设备：MSR5660
版本：R0305P08



PC通过二层网络连接到MSR5660，网关为MSR5660的G0/0接口。在PC上tracert Server，发现tracert回显异常。
正常回显，可以正确显示下一跳。
C:\Documents and Settings\Administrator>tracert -d 192.168.2.1
Tracing route to 192.168.2.1 over a maximum of 30 hops
1 <1 ms <1 ms <1 ms 192.168.1.254
2 1 ms 1 ms <1 ms 10.1.12.2
3 2 ms 1 ms 1 ms 10.1.23.3
4 2 ms 2 ms 2 ms 192.168.2.1
Trace complete.
异常回显，除最后一跳，下一跳都为网关地址。
C:\Documents and Settings\Administrator>tracert -d 192.168.2.1
Tracing route to 192.168.2.1 over a maximum of 30 hops
1 <1 ms <1 ms <1 ms **192.168.1.254**
2 1 ms 1 ms 1 ms **192.168.1.254**
3 2 ms 1 ms 1 ms **192.168.1.254**
4 2 ms 2 ms 3 ms 192.168.2.1
Trace complete.
通过测试发现只有最后一跳可以正确显示，前面所有回显都为网关地址。在PC上和R2与R3之间抓包分析。

```
24 14:37:26.846077   192.168.1.1     192.168.2.1     ICMP   106 Echo (ping) request  id=0x0200, seq=10240/40, ttl=3 (no response found!)
25 14:37:26.848344   192.168.1.254   192.168.1.1     ICMP    70 Time-to-live exceeded (Time to live exceeded in transit)
```

```
88 08:23:12.564006   192.168.1.1     192.168.2.1     ICMP   106 Echo (ping) request  id=0x0200, seq=10240/40, ttl=1 (no response found!)
89 08:23:12.564342   10.1.23.3       192.168.1.1     ICMP    70 Time-to-live exceeded (Time to live exceeded in transit)
```

通过抓包分析，R3已经正确回复了ICMP差错报文（TTL超时），源IP地址为R3的接口地址。经过R1后，源地址变为了R1的接口地址。也就是MSR5660将ICMP差错报文的源地址进行了更改。

```
33 14:37:27.846433   192.168.1.1     192.168.2.1     ICMP   106 Echo (ping) request  id=0x0200, seq=11008/43, ttl=4
34 14:37:27.849450   192.168.2.1     192.168.1.1     ICMP   106 Echo (ping) reply    id=0x0200, seq=11008/43, ttl=252
```

```
102 08:23:13.564467   192.168.1.1     192.168.2.1     ICMP   106 Echo (ping) request  id=0x0200, seq=11008/43, ttl=2
103 08:23:13.565701   192.168.2.1     192.168.1.1     ICMP   106 Echo (ping) reply    id=0x0200, seq=11008/43, ttl=254
```

通过抓包分析，发送至最后一跳，也就是tracert的目的设备的报文没有被转换，该报文为正常的ICMP Echo reply报文，非ICMP差错报文。

经确认，我司产品实现原理参照RFC5508：

address varies depending on whether the Basic NAT or NAPT function [NAT-TRAD] is enforced by the NAT device. A NAT device enforcing the Basic NAT function has a pool of public IP addresses and enforces address mapping (which is different from the endpoint mapping enforced by NAPT) when a private node initiates an outgoing session via the NAT device. So, if the NAT device has active mapping for the IP address of the intermediate node Router-y, the NAT device MUST translate the source IP address of the ICMP Error packet with the public IP address in the mapping. In all other cases, the NAT device MUST simply use its own IP address in the external domain to translate the source IP address.

从RFC来看，当有nat session时，使用public地址转换ICMP差错报文源地址，其它情况使用接口地址转换ICMP差错报文源地址。其目的是为了保护私网地址不被泄露。

检查设备配置，发现内网接口有不相关的NAT配置，当从内网PC Tracert到公网的时候，回程的ICMP差错报文没有匹配上nat server，使用地址转换ICMP差错报文源地址，导致所有的差错报文都为G0/0的接口地址，最后一跳由于是正常的ICMP Echo reply报文，该报文没有被转换。PC上Tracert看到的现象就是除了最后一跳，剩下的每一跳都为网关G0/0的接口地址。
[H3C-GigabitEthernet0/0]display this

#
interface GigabitEthernet0/0
 port link-mode route
 combo enable copper
 ip address 192.168.1.254 255.255.255.0
 ospf 1 area 0.0.0.0
 nat server protocol tcp global 1.1.1.1 inside 2.2.2.2
#
由于客户内网并无NAT Server需求，将G0/0接口下的NAT Server去掉后，Tracert回显正常。
已提交需求，针对RFC5508功能做一个开关控制。