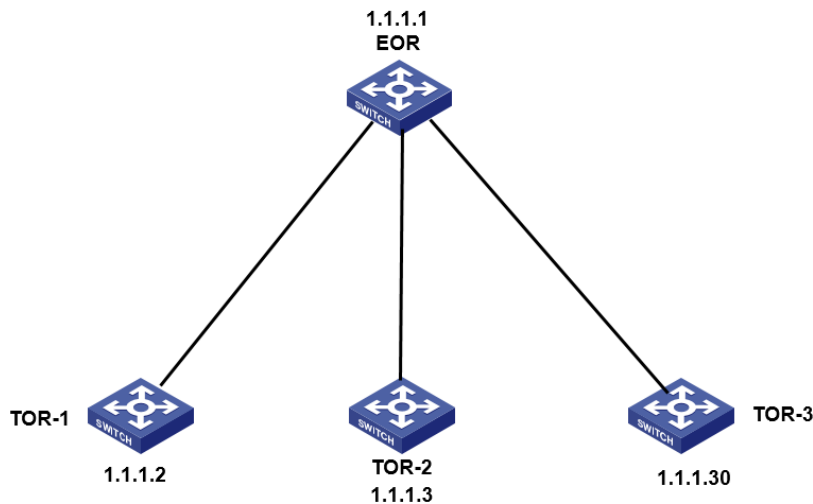


知 某局点 S6520-54QF-EI ping 延迟抖动过大

IGMP 刘雨 2020-04-13 发表

组网及说明

现场组网图如下，所有设备均为两台S6520-54QF-EI设备堆叠，四台设备之间的链路均为跨框的聚合链路，版本为Version 7.1.045, Release 2432P05。



问题描述

1. 上层EOR交换机ping下层所有TOR交换机报文抖动比较严重，结果如下：

```
Ping 1.1.1.3 (1.1.1.3): 56 data bytes, press CTRL_C to break
56 bytes from 1.1.1.3: icmp_seq=0 ttl=255 time=57.380 ms
56 bytes from 1.1.1.3: icmp_seq=1 ttl=255 time=2.503 ms
56 bytes from 1.1.1.3: icmp_seq=2 ttl=255 time=2.519 ms
56 bytes from 1.1.1.3: icmp_seq=3 ttl=255 time=2.301 ms
56 bytes from 1.1.1.3: icmp_seq=4 ttl=255 time=42.491 ms
56 bytes from 1.1.1.3: icmp_seq=5 ttl=255 time=125.988 ms
56 bytes from 1.1.1.3: icmp_seq=6 ttl=255 time=2.705 ms
56 bytes from 1.1.1.3: icmp_seq=7 ttl=255 time=2.431 ms
56 bytes from 1.1.1.3: icmp_seq=8 ttl=255 time=2.547 ms
56 bytes from 1.1.1.3: icmp_seq=9 ttl=255 time=2.497 ms
56 bytes from 1.1.1.3: icmp_seq=10 ttl=255 time=148.613 ms
56 bytes from 1.1.1.3: icmp_seq=11 ttl=255 time=2.998 ms
56 bytes from 1.1.1.3: icmp_seq=12 ttl=255 time=2.325 ms
56 bytes from 1.1.1.3: icmp_seq=13 ttl=255 time=2.518 ms
56 bytes from 1.1.1.3: icmp_seq=14 ttl=255 time=2.278 ms
56 bytes from 1.1.1.3: icmp_seq=15 ttl=255 time=149.082 ms
56 bytes from 1.1.1.3: icmp_seq=16 ttl=255 time=2.608 ms
56 bytes from 1.1.1.3: icmp_seq=17 ttl=255 time=2.923 ms
56 bytes from 1.1.1.3: icmp_seq=18 ttl=255 time=2.272 ms
56 bytes from 1.1.1.3: icmp_seq=19 ttl=255 time=2.649 ms
56 bytes from 1.1.1.3: icmp_seq=20 ttl=255 time=153.445 ms
--- Ping statistics for 1.1.1.2 ---
20 packet(s) transmitted, 20 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 2.149/16.580/149.156/35.903 ms

--- Ping statistics for 1.1.1.3 ---
20 packet(s) transmitted, 20 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.877/26.721/128.548/37.844 ms

--- Ping statistics for 1.1.1.30 ---
20 packet(s) transmitted, 20 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.958/21.987/151.206/47.417 ms
2、下层TOR交换机分别ping EOR交换机结果正常：
--- Ping statistics for 1.1.1.1 ---
```

20 packet(s) transmitted, 20 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 0.230/0.284/0.341/0.031 ms

--- Ping statistics for 1.1.1.1 ---

20 packet(s) transmitted, 20 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.722/1.959/2.882/0.248 ms

--- Ping statistics for 1.1.1.1 ---

20 packet(s) transmitted, 20 packet(s) received, 0.0% packet loss
round-trip min/avg/max/std-dev = 1.680/2.104/3.339/0.312 ms

过程分析

(1) 打开debug rtx, 从EOR ping TOR-2, 可以发现发往1.1.1.3的ping包从1框CPU发出, 然后从1框的1/0/5口发出(1框为主框), 发送到对端TOR-2的2/0/47口:

[EOR-1-probe]ping 1.1.1.3

Ping 1.1.1.3 (1.1.1.3): 56 data bytes, press CTRL_C to break

*Jan 19 06:59:56:659 2011 EOR-1 DRVPLAT/7/RxTxDebug:

From board 1: transmit packet from chip0,port5,Priority=3,len=102

0000 70 57 bf 30 10 f9 7c 1e 06 2c 1f 9a 81 00 00 c8
0010 08 00 45 00 00 54 fc 50 00 00 ff 01 bb 52 01 01
0020 01 01 01 01 01 03 08 00 01 17 25 dc 00 00 4d 36
0030 8b ec 00 0a 0c dd 08 09 0a 0b 0c 0d 0e 0f 10 11

56 bytes from 1.1.1.3: icmp_seq=0 ttl=255 time=133.296 ms

56 bytes from 1.1.1.3: icmp_seq=1 ttl=255 time=2.477 ms

*Jan 19 06:59:56:993 2011 EOR-1 DRVPLAT/7/RxTxDebug:

From board 1: transmit packet from chip0,port5,Priority=3,len=102

0000 70 57 bf 30 10 f9 7c 1e 06 2c 1f 9a 81 00 00 c8
0010 08 00 45 00 00 54 fc 51 00 00 ff 01 bb 51 01 01
0020 01 01 01 01 01 03 08 00 e8 7b 25 dc 00 01 4d 36
0030 8b ec 00 0f 25 72 08 09 0a 0b 0c 0d 0e 0f 10 11

(2) 从对端TOR-2 ping回来的reply报文从2/0/45口发出, 到EPR备框的2/0/8口:

[EOR-1-probe]ping 1.1.1.3

Ping 1.1.1.3 (1.1.1.3): 56 data bytes, press CTRL_C to break

56 bytes from 1.1.1.3: icmp_seq=0 ttl=255 time=3.220 ms

*Jan 19 07:03:29:505 2011 EOR-1 DRVPLAT/7/RxTxDebug: -Slot=2;

From board 2: received packet from

chip0,port8,reason=0x0,cos=9,sMod=6,sPort=8,len=102,Matched=0,time=0,src_vp=-1

*Jan 19 07:03:29:505 2011 EOR-1 DRVPLAT/7/RxTxDebug: -Slot=2;

0000 7c 1e 06 2c 1f 9a 70 57 bf 30 10 f9 81 00 00 c8
0010 08 00 45 00 00 54 02 74 00 00 ff 01 b5 2f 01 01
0020 01 03 01 01 01 01 00 00 f6 26 25 fe 00 00 4d 36
0030 8c c1 00 0a 1e d6 08 09 0a 0b 0c 0d 0e 0f 10 11

56 bytes from 1.1.1.3: icmp_seq=1 ttl=255 time=22.830 ms

(3) 堆叠场景下, 从备板上来的ping包, 先送往备板的CPU, 然后备板CPU再走ipc通道, 通过hg口发过主板CPU。这时候Ping包和堆叠两框间的交互的协议报文、同步报文等走的是同样的队列, 同时多了一个CPU软转的过程, 而且icmp的优先级相对其他交互报文来说较低, 所以会出现偶尔的跨版ping延迟大, 属于正常现象。这个不会影响跨框的流量转发, 因为正常的业务流量走的是硬件转发流程, 从备板口上来之后, hg口到主板然后直接硬件表项转发, 不会上CPU。

解决方法

该现象属于堆叠状态下正常现象, 不会影响跨框的流量转发, 因为正常的业务流量走的是硬件转发流程, 从备板口上来之后, 从hg口到主板然后直接硬件表项转发, 不会上CPU。