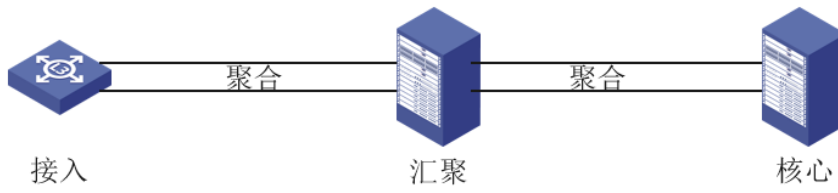


# 某局点聚合口HASH不均匀的经验案例

二层链路聚合 张腾 2021-01-25 发表

## 组网及说明



拓扑说明：汇聚是2台设备做IRF，上行和下行是跨框链路聚合

## 问题描述

汇聚下行聚合口HASH不均匀

## 过程分析

1、汇聚下行聚合口成员接口流量情况如下，可以看到流量相差很大

Ten-GigabitEthernet1/0/0/14

Last 300 second input: 17408 packets/sec 5984513 bytes/sec 0%

Last 300 second output: 53550 packets/sec 67264446 bytes/sec 5%

Ten-GigabitEthernet2/0/0/14

Last 300 second input: 17121 packets/sec 5290434 bytes/sec 0%

Last 300 second output: 469 packets/sec 551933 bytes/sec 0%

2、检查配置未修改HASH参数，即采用缺省的HASH方法，转发二层数据帧时，采用的聚合负载分担类型包括以太网类型、源MAC地址、目的MAC地址和源服务端口。怀疑现场业务流量五元组相同，是同一条流导致无法负载，于是在聚合口所有成员接口抓包查看，发现流量大多为组播流，并且业务流量类型也较为丰富，不应该出现成员接口流量相差很大的情况。

| 源地址                     | 源端口   | 目的地址         | 目的端口 | 分组    | 包/秒     | 平均带宽 (bps) | 最大带宽 (bps) | 最大突发        | 突发警报 | 最大缓冲区 (B) | 缓冲区警报 |
|-------------------------|-------|--------------|------|-------|---------|------------|------------|-------------|------|-----------|-------|
| fe80::4139:53c1fa7e7e3d | 63325 | m02:1:3      | 5355 | 2     | 19.64   | 13 k       | 0          | 1 / 100ms   | 0    | 86        | 0     |
| fe80::4139:53c1fa7e7e3d | 546   | m02:1:2      | 547  | 1     | 0.00    | 0          | 0          | 1 / 100ms   | 0    | 157       | 0     |
| fe80::4139:53c1fa7e7e3d | 54968 | m02:1:3      | 5355 | 2     | 19.91   | 13 k       | 0          | 1 / 100ms   | 0    | 86        | 0     |
| fe80::4139:53c1fa7e7e3d | 53936 | m02:1:3      | 5355 | 2     | 20.09   | 13 k       | 0          | 1 / 100ms   | 0    | 86        | 0     |
| fe80::4139:53c1fa7e7e3d | 51842 | m02:1:3      | 5355 | 2     | 20.10   | 13 k       | 0          | 1 / 100ms   | 0    | 86        | 0     |
| 192.168.23.21           | 4000  | 230.1.0.126  | 1001 | 25988 | 2361.79 | 25 M       | 26 M       | 241 / 100ms | 1    | 28 M      | 1     |
| 192.168.23.21           | 4000  | 230.1.0.49   | 1001 | 6257  | 568.74  | 6178 k     | 6192 k     | 57 / 100ms  | 1    | 1624 k    | 1     |
| 192.168.23.13           | 4000  | 230.1.0.101  | 1001 | 39511 | 3590.81 | 39 M       | 39 M       | 367 / 100ms | 1    | 46 M      | 1     |
| 192.168.23.13           | 4000  | 230.1.9.1    | 1001 | 10392 | 944.51  | 10 M       | 10 M       | 96 / 100ms  | 1    | 7241 k    | 1     |
| 192.168.23.13           | 4000  | 230.1.0.134  | 1001 | 39478 | 3587.82 | 38 M       | 39 M       | 366 / 100ms | 1    | 46 M      | 1     |
| 192.168.23.13           | 4000  | 230.1.0.58   | 1001 | 6235  | 566.74  | 6157 k     | 6301 k     | 58 / 100ms  | 2    | 1594 k    | 1     |
| 192.168.21.15           | 4000  | 230.1.0.56   | 1001 | 40326 | 3664.89 | 39 M       | 40 M       | 375 / 100ms | 1    | 47 M      | 1     |
| 192.168.21.15           | 4000  | 230.1.0.143  | 1001 | 33245 | 3021.43 | 32 M       | 33 M       | 309 / 100ms | 1    | 38 M      | 1     |
| 192.168.10.14           | 0     | 230.1.0.52   | 1001 | 4961  | 451.01  | 4899 k     | 6409 k     | 59 / 100ms  | 246  | 40 k      | 257   |
| 169.254.126.61          | 55579 | 224.0.0.255  | 1900 | 1     | 0.00    | 0          | 0          | 1 / 100ms   | 0    | 175       | 0     |
| 169.254.126.61          | 58436 | 224.0.0.252  | 5355 | 2     | 19.89   | 10 k       | 0          | 1 / 100ms   | 0    | 66        | 0     |
| 169.254.126.61          | 49478 | 224.0.0.252  | 5355 | 2     | 19.95   | 10 k       | 0          | 1 / 100ms   | 0    | 66        | 0     |
| 169.254.126.61          | 63159 | 224.0.0.252  | 5355 | 2     | 20.12   | 10 k       | 0          | 1 / 100ms   | 0    | 66        | 0     |
| 169.254.126.61          | 51503 | 224.0.0.252  | 5355 | 2     | 20.12   | 10 k       | 0          | 1 / 100ms   | 0    | 66        | 0     |
| 10.222.4.187            | 2007  | 230.1.8.8    | 1001 | 10390 | 944.32  | 10 M       | 10 M       | 96 / 100ms  | 1    | 7239 k    | 1     |
| 10.222.4.181            | 2007  | 230.1.8.3    | 1001 | 10425 | 947.51  | 10 M       | 10 M       | 96 / 100ms  | 1    | 7286 k    | 1     |
| 10.222.4.175            | 2007  | 230.1.8.6    | 1001 | 10426 | 947.60  | 10 M       | 10 M       | 96 / 100ms  | 1    | 7287 k    | 1     |
| 10.222.2.167            | 2563  | 230.1.8.5    | 1001 | 9354  | 850.25  | 9237 k     | 9343 k     | 86 / 100ms  | 1    | 5831 k    | 1     |
| 10.222.3.121            | 2567  | 230.1.8.1    | 1001 | 10434 | 948.33  | 10 M       | 10 M       | 96 / 100ms  | 1    | 7298 k    | 1     |
| 10.222.3.114            | 4000  | 230.1.10.1   | 1001 | 8943  | 812.80  | 8830 k     | 9017 k     | 83 / 100ms  | 1    | 5273 k    | 1     |
| 10.222.2.190            | 2561  | 230.1.9.52   | 1001 | 8941  | 812.58  | 8827 k     | 9125 k     | 84 / 100ms  | 1    | 5270 k    | 1     |
| 10.222.2.150            | 4000  | 230.1.10.5   | 1001 | 8920  | 810.75  | 8808 k     | 8908 k     | 82 / 100ms  | 1    | 5242 k    | 1     |
| 10.222.2.150            | 4000  | 230.1.10.3   | 1001 | 7000  | 636.25  | 6912 k     | 6952 k     | 64 / 100ms  | 1    | 2633 k    | 1     |
| 10.222.2.150            | 4000  | 230.1.10.8   | 1001 | 7005  | 636.70  | 6917 k     | 6952 k     | 64 / 100ms  | 1    | 2640 k    | 1     |
| 10.222.2.150            | 4000  | 230.1.10.6   | 1001 | 7002  | 636.43  | 6914 k     | 6952 k     | 64 / 100ms  | 1    | 2636 k    | 1     |
| 10.222.2.120            | 49299 | 239.2.15.103 | 8005 | 2082  | 189.34  | 2056 k     | 2064 k     | 19 / 100ms  | 0    | 2241      | 0     |
| 10.222.2.120            | 59304 | 239.2.15.106 | 8005 | 2082  | 189.25  | 2055 k     | 2064 k     | 19 / 100ms  | 0    | 2213      | 0     |
| 10.222.2.120            | 56251 | 239.2.15.105 | 8005 | 2076  | 188.81  | 2051 k     | 2064 k     | 19 / 100ms  | 0    | 2205      | 0     |
| 10.222.2.120            | 58135 | 239.2.15.109 | 8005 | 2078  | 188.99  | 2053 k     | 2064 k     | 19 / 100ms  | 0    | 2193      | 0     |
| 10.222.2.120            | 59016 | 239.2.15.112 | 8005 | 2071  | 188.35  | 2046 k     | 2064 k     | 19 / 100ms  | 0    | 2193      | 0     |
| 10.222.2.120            | 49467 | 239.2.15.101 | 8005 | 2081  | 189.26  | 2056 k     | 2064 k     | 19 / 100ms  | 0    | 1627      | 0     |

3、汇聚设备做了IRF，考虑到设备默认是本地优先转发即流量从哪台设备进入就优先从那台设备转发出去，不会负载分担；于是将汇聚设备的本地优先转发功能关闭观察效果，发现故障依旧。

取消聚合负载分担采用本地转发优先  
system-view

[Sysname] undo link-aggregation load-sharing mode local-first

5、根据现场拓扑发现存在二次HASH的情况，即流量从核心到达汇聚HASH一次，汇聚再发给接入又HASH一次；并且核心和汇聚都采用缺省的HASH方式，即2次HASH的方式相同。经确认如果是经过二次hash的流量并且HASH方式相同，基本很难hash开，可通过调整和第一次HASH不同的策略观察效果。

## 解决方法

修改汇聚设备的HASH因子:

#配置二层聚合接口1对应的聚合组内按照报文目的源IP、源端口、目的IP、目的端口进行聚合负载分担。

```
system-view
```

```
[Sysname] interface bridge-aggregation 1
```

```
[Sysname-Bridge-Aggregation1] link-aggregation load-sharing mode destination-port destination-ip s  
ource-ip source-port
```

