

知 某局点 S125G-AF ECMP流量哈希不均问题

配置优化 倪民 2021-12-14 发表

组网及说明

不涉及

问题描述

客户两台交换机做了堆叠，与下行四台服务器直连，每框各出一根线做二层聚合连接，建立EBGP ipv4和ipv6邻居，四台服务器宣告同一网段，在设备侧能学习到4条等价路由，出接口分别是BAGG7、11、15、19。但查看4个下一跳聚合口的300s流量大小，发现聚合组7和15，接口流量较大，11和19相比之下就小很多。

```
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>dis interface Bridge-Aggregation7 | in %
Last 300 seconds input: 26010 packets/sec 15739450 bytes/sec 0%
Last 300 seconds output: 23762 packets/sec 35478654 bytes/sec 0%
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>dis interface Bridge-Aggregation11 | in %
Last 300 seconds input: 349 packets/sec 373380 bytes/sec 0%
Last 300 seconds output: 552 packets/sec 733195 bytes/sec 0%
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>dis interface Bridge-Aggregation15 | in %
Last 300 seconds input: 25781 packets/sec 16420656 bytes/sec 0%
Last 300 seconds output: 22247 packets/sec 33097108 bytes/sec 0%
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>dis interface Bridge-Aggregation19 | in %
Last 300 seconds input: 555 packets/sec 737726 bytes/sec 0%
Last 300 seconds output: 459 packets/sec 553118 bytes/sec 0%
```

过程分析

- 1、现场打的流量也是从IRF的两个框分别进入，不涉及本框优转；
- 2、现场将打流设备直接连接到交换机，测试哈希正常，负载均匀。

```
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>dis interface Bridge-Aggregation7 | in %
Last 300 seconds input: 343310 packets/sec 432934245 bytes/sec 8%
Last 300 seconds output: 404451 packets/sec 449586366 bytes/sec 9%
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>dis interface Bridge-Aggregation11 | in %
Last 300 seconds input: 317559 packets/sec 418216002 bytes/sec 8%
Last 300 seconds output: 380654 packets/sec 406245872 bytes/sec 8%
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>dis interface Bridge-Aggregation15 | in %
Last 300 seconds input: 332544 packets/sec 414490728 bytes/sec 8%
Last 300 seconds output: 316417 packets/sec 449077101 bytes/sec 9%
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>dis interface Bridge-Aggregation19 | in %
Last 300 seconds input: 304946 packets/sec 397989721 bytes/sec 8%
Last 300 seconds output: 380753 packets/sec 415902218 bytes/sec 8%
<HEXA-SJ5-Q109_Q117-CSW-H12508-12U08_12U19>
```

解决方法

- 1、经确认，S125G-AF还有上行设备，流量在进入S125G之前已经被哈希过一次。
- 2、如果之前被hash过，例如TCP端口号奇数都从1/4/0/1进，偶数从2/4/0/1进，设备上再次采用和上行设备相同的哈希算法，就会导致二次哈希极化，引发现场的情况。
- 3、设备上修改参与计算的哈希因子。不要用默认的基于源IP地址、目的IP地址、源端口和目的端口
 - 配置负载分担方式，目前仅支持以下四种组合方式：
 - 基于源IP地址
 - 基于源IP地址和目的IP地址
 - 基于源IP地址、源端口和目的端口
 - 基于源IP地址、目的IP地址、源端口和目的端口
- 4、现场修改哈希因子后解决

