6127XLG LACP不同配置顺序导致结果不同，以至于HPE刀片服务器系统下网卡绑定不成功

LACP　赵晓静　2017-11-03 发表

硬件信息：

1、HPE C7000 刀箱

2、6127XLG交换机（2台6127XLG在interconnect bay3和bay4槽位）

3、HPE ProLiant BL460c Gen9（16台刀片的Mezz slot1上均是560M的网卡）

操作系统：

Ubuntu 14.04

问题现象：

用户在系统下对MeZZanine1:1和MeZZanine1:2对应的560M两个网口配置了基于二层哈希的网卡绑定，与之port mapping的两个IRF的6127XLG配置动态LACP，协商报文一直只有ACT，并且没有AGG进一步的状态变更，而更换成静态LACP，交换机侧不再发协商报文。且两种LACP模式的下联口聚合端口的状态都是unselected，表示未协商成功。导致系统下网卡负载均衡绑定不成功。

Ubuntu操作系统下报错：

[Port 2: rx_machine] selected <- UNSELECTED

[Port 2: rx_machine] Slave 2,partner state changed!

[Port 2: mux_machine] ->DETACHED

[Port 1: mux_machine] DETACHED -> WAITING

[Port 2: mux_machine] DETACHED -> WAITING

[Port 1: mux_machine] ATTACHED Entered

[Port 2: mux_machine] ATTACHED Entered

[Port 1: mux_machine] ATTACHED -> COLLECTING

[Port 2: mux_machine] ATTACHED -> COLLECTING

6127XLG交换机测显示异常，两个LACP端口全部Unselected

AGG Interface	AGG Mode	Partner ID	Selected Ports	Unselected Ports	Individual Ports	Share Type
BAGG1	S	None	0	2	0	Shar
BAGG2	S	None	0	2	0	Shar

一、6127XLG交换机静态LACP：

聚合配置如下：

interface Bridge-Aggregation2

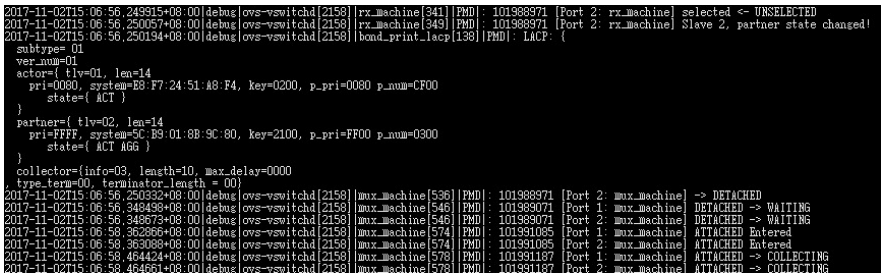
port link-type trunk

undo port trunk permit vlan 1

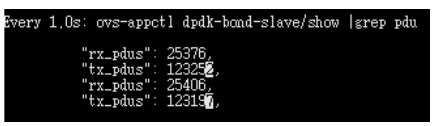
port trunk permit vlan 112 to 120

undo stp enable

1.服务器内部协商状态如下，此时服务器侧收到的协商报文，交换板侧状态只有ACT，没有AGG。



2.服务器侧的收发协商报文，没有再收到交换机侧的报文，未协商成功。



3.此时交换机状态，聚合端口全部unselected：

```
[HPE]display link-aggregation s
ggregation Interface Type:
BAGG -- Bridge-Aggregation, BLAGG -- Blade-Aggregation, RAGG -- Route-Aggregation
ggregation Mode: S -- Static, D -- Dynamic
oadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
ctor System ID: 0x8000, e8f7-2451-a8f4

GG         AGG   Partner ID             Selected  Unselected  Individual  Share
nterface   Mode                         Ports     Ports       Ports       Type
--------------------------------------------------------------------------------
BAGG1      D     0x8000, 0000-0000-0000  0         2           0           Shar
BAGG2      S     None                    0         2           0           Shar
BAGG100    D     0x8000, f02f-a73b-0101  2         0           0           Shar
[HPE]
```

二、6127XLG交换机动态LACP：

聚合配置如下：

interface Bridge-Aggregation2

port link-type trunk

undo port trunk permit vlan 1

port trunk permit vlan 112 to 120

link-aggregation mode dynamic

undo stp enable

1.服务器内部协商状态如下，此时服务器侧收到的协商报文，交换板侧状态只有ACT，没有AGG。

```
2017-11-02T15:29:51.849014+08:00|debug|ovs-vswitchd[2158]|rx_machine[341]|PMD|: 103364560 [Port 2: rx_machine] selected <- UNSELECTED
2017-11-02T15:29:51.849215+08:00|debug|ovs-vswitchd[2158]|rx_machine[349]|PMD|: 103364560 [Port 2: rx_machine] Slave 2, partner state changed!
2017-11-02T15:29:51.849401+08:00|debug|ovs-vswitchd[2158]|bond_print_lacp[138]|PMD|: LACP: {
  subtype= 01
  ver_num=01
  actor={ tlv=01, len=14
    pri=0080, system=E8:F7:24:51:A8:F4, key=0200, p_pri=0080 p_num=CF00
      state={ ACT }
  }
  partner={ tlv=02, len=14
    pri=FFFF, system=5C:B9:01:8B:9C:80, key=2100, p_pri=FF00 p_num=0300
      state={ ACT AGG }
  }
  collector={info=03, length=10, max_delay=0000
  type_term=00, terminator_length = 00}
2017-11-02T15:29:51.849804+08:00|debug|ovs-vswitchd[2158]|mux_machine[536]|PMD|: 103364560 [Port 2: mux_machine] -> DETACHED
2017-11-02T15:29:51.949512+08:00|debug|ovs-vswitchd[2158]|rx_machine[341]|PMD|: 103364660 [Port 2: rx_machine] selected <- UNSELECTED
2017-11-02T15:29:51.949730+08:00|debug|ovs-vswitchd[2158]|rx_machine[349]|PMD|: 103364660 [Port 2: rx_machine] Slave 2, partner state changed!
2017-11-02T15:29:51.949923+08:00|debug|ovs-vswitchd[2158]|bond_print_lacp[138]|PMD|: LACP: {
  subtype= 01
  ver_num=01
  actor={ tlv=01, len=14
    pri=0080, system=E8:F7:24:51:A8:F4, key=0200, p_pri=0080 p_num=CF00
      state={ ACT }
  }
  partner={ tlv=02, len=14
    pri=FFFF, system=5C:B9:01:8B:9C:80, key=2100, p_pri=FF00 p_num=0300
      state={ ACT AGG }
  }
  collector={info=03, length=10, max_delay=0000
  type_term=00, terminator_length = 00}
2017-11-02T15:29:52.049446+08:00|debug|ovs-vswitchd[2158]|mux_machine[546]|PMD|: 103364760 [Port 2: mux_machine] DETACHED -> WAITING
```

2.此时服务器侧和交换机一直在协商状态，交换机侧也会一直发送协商报文，未协商成功。

```
Every 1.0s: ovs-appctl dpdk-bond-slave/show |grep pdu

          "rx_pdus": 25752,
          "tx_pdus": 123670,
          "rx_pdus": 25778,
          "tx_pdus": 123612,
```

3.此时交换机状态unselected：

```
[HPE]display link-aggregation summary
ggregation Interface Type:
BAGG -- Bridge-Aggregation, BLAGG -- Blade-Aggregation, RAGG -- Route-Aggregation
ggregation Mode: S -- Static, D -- Dynamic
oadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
ctor System ID: 0x8000, e8f7-2451-a8f4

GG         AGG   Partner ID             Selected  Unselected  Individual  Share
nterface   Mode                         Ports     Ports       Ports       Type
--------------------------------------------------------------------------------
BAGG1      D     0x8000, 0000-0000-0000  0         2           0           Shar
BAGG2      D     0x8000, 0000-0000-0000  0         2           0           Shar
BAGG100    D     0x8000, f02f-a73b-0101  2         0           0           Shar
[HPE]
```

用户一共16把刀片，由于配置的都是IRF后6127XLG对应的下联口，于是远程连接用户环境，选择一台新刀片配置，发现按照如下顺序配置，则6127XLG链路聚合端口状态全部为selected，并且操作系统下网卡绑定也配置成功。

[HPE]interface Bridge-Aggregation 1

[HPE-Bridge-Aggregation1]link-aggregation mode dynamic

[HPE-Bridge-Aggregation1]quit

[HPE]interface TwentyGigE 1/0/6

[HPE-TwentyGigE1/0/6]port link-aggregation group 1

[HPE-TwentyGigE1/0/6]quit

[HPE]interface TwentyGigE 2/0/6

[HPE-TwentyGigE2/0/6]port link-aggregation group 1

[HPE]interface Bridge-Aggregation 1

[HPE-Bridge-Aggregation1]port link-type trunk

Configuring TwentyGigE1/0/6 done.

Configuring TwentyGigE2/0/6 done.

[HPE-Bridge-Aggregation1]port trunk permit vlan 112 to 120

Configuring TwentyGigE1/0/6 done.

Configuring TwentyGigE2/0/6 done.

[HPE-Bridge-Aggregation1]save
[HPE]display link-aggregation summary
Aggregation Interface Type:
BAGG -- Bridge-Aggregation, BLAGG -- Blade-Aggregation, RAGG -- Route-Aggregation
Aggregation Mode: S -- Static, D -- Dynamic
Loadsharing Type: Shar -- Loadsharing, NonS -- Non-Loadsharing
Actor System ID: 0x8000, 40b9-3ca0-1a19

AGG Interface	AGG Mode	Partner ID	Selected Ports	Unselected Ports	Individual Ports	Share Type
BAGG1	S	None	2	0	0	Shar

6127XLG交换机与6125系列配置基本一样，所以此结论也适用于6125系列交换机，LACP不同配置顺序导致结果不同，以至于会导致系统下网卡绑定不成功。所以必须按照如下顺序配置，否则会导致链路聚合端口状态Unselected。
先创建链路聚合组—>设置LACP Mode—>端口加入聚合组—>聚合组接口配置link-type，vlan等其余配置