CPE

Layer-2 Acceleration Support

Application Note

Table of Contents

1 Introduction

This document describes packet handling when a CPE is configured in Layer-2 (L2) acceleration mode.

1.1 Background

The existing Layer-3 packet acceleration implementation is based on 5-tuples, which is basically Layer-3 (IPSA, IPDA, protocol), and Layer-4 fields (TCP/UDP source and destination ports). A L3 flow is uniquely identified by a combination of the various fields of the 5-tuple. If a packet matches the 5-tuples of a learned flow, it is considered a flow hit, and the packet is accelerated. If a packet does not match any of the existing learned flows, it is considered a flow miss and the next higher level accelerator tries to accelerate the packet. If the packet is a flow miss in the next level accelerator (flow cache) also then the packet may be blogged and sent to Linux networking stack. It is possible that this packet when going out of an egress interface is learned by flow cache, and is finally learned by the hardware accelerator.

1.2 Existing Limitations

The L3 acceleration works well for routed and NAT'ed flows, but has the following restrictions/limitations for bridging (L2 acceleration):

- L3 acceleration flows are too fine for Layer-2 acceleration.
  Layer-3 acceleration creates too many flows for a single L2 flow (when the only fields that are changing are L3/4 fields).
- L3 acceleration supports only IP protocol with TCP/UDP, a few L3 test protocols, and tunneling protocols such as, 6RD, DS-LITE, GRE, and L2TP.
  This is a big limitation for L2, that requires all L3/4 to be accelerated.
- L3 acceleration ignores the VLAN tags in unicast flows.
  VLAN tags are essential for L2 acceleration.

These restrictions when using L3 acceleration in bridging mode, are the reasons for implementing L2 acceleration.

1.3 Layer-2 Flow Tuple

The flows in L2 acceleration are uniquely identified by a tuple based on L2 fields and IP-ToS value.

NOTE: L2 Flow Acceleration is only applicable for unicast traffic.

The L2-tuple consists of:

- Destination MAC
- Source MAC
- Ether Type
- Outer and inner VLAN Tags
- Number of VLAN Tags 0, 1 or 2
- IP-ToS (8-bits)
- Ingress interface
- GEM flow for PON

All the above fields are available when a packet is received (except the VLAN tags in some cases). When a VLAN tag is not present in a packet a default value (say 0xFFFFFFFF) is used.

The current implementation supports a maximum of two VLAN tags.

- When an untagged packet is received, the number of VLAN tags is 0, both VLAN tag0 and VLAN tag1 are set to fixed value (0xFFFFFFFF or 0x0).
- When a single tag packet is received, the number of VLAN tags is 1, VLAN tag0 is set to received tag, and VLAN tag1 is set to fixed value (0xFFFFFFFF or 0x0).
- When a double tag packet is received, the number of VLAN tags is 2, VLAN tag0 is set to received outer tag, and VLAN tag1 is set to received inner tag.

2 Packet Acceleration Mode

Broadcom CPE supports two packet acceleration modes:

- **L3 acceleration**: In this acceleration mode, a packet is accelerated based on L3+L4 tuple even when the CPE is configured as a bridge. If packet L3 info is non-IP, it is not accelerated.
- **L2 and L3 acceleration**: In this acceleration mode, a packet is accelerated based on the destination MAC address (MAC DA). If the MAC DA in the received packet matches one of the configured MAC on the CPE, the packet is treated as routed/NAT'd and L3 acceleration is used. Otherwise, the packet is treated as bridged and L2 acceleration is used. In the L2 acceleration mode even non-IP packets are accelerated.
  This is the default mode when the system comes up.

2.1 CLI Commands

2.1.1 Display Acceleration Mode

The current acceleration mode is displayed by using the following CLI command:

```
# fc status
```

The output is similar to this:

```
Flow Timer Interval<0xbf1a686c> = 10000 millisecs
Pkt-HW Activate Deferral<0xbf1a6a64> : 1
Pkt-HW Idle Deactivate<0xbf1a6a68> = 0
Acceleration Mode: <L2 & L3>
MCast Learning <Disabled>
MCast Acceleration IPv4<Enabled> IPv6<Enabled>
IPv6 Learning <Enabled>
GRE Learning <Enabled> Mode<Tunnel>
Flow Learning Enabled : Max<16384>, Active<0>, Cummulative [ 4 - 4 ]
```

2.1.2 Change Acceleration Mode

Change the acceleration mode by using the following CLI command:

```
# fc config --accel-mode m
```

where,

`m` = 0 is L3 acceleration mode

`m` = 1 is L2 & L3 acceleration mode

2.2 L2 Acceleration Support

2.2.1 Software Acceleration

All Broadcom CPEs will support L2 acceleration in software using flow cache module.

2.2.2 Hardware Acceleration

L2 acceleration in hardware is currently supported by Runner-based platforms: BCM63138, BCM63148, BCM6858X, BCM68560, BCM68360, BCM6846X, with exception of BCM6838X and BCM6848X. It is not supported by non-Runner based platforms such as BCM63268.

Note: On platforms where L2 acceleration is not supported by hardware accelerator, if L2 acceleration mode is configured, L2 flows are accelerated by software only, but L3 can be accelerated by both software and hardware accelerators.

3 Flows

Broadcom CPE parses the received packets, and classifies them based on the tuple, and assigns them to the flows. All the packets having the same tuple field values are classified as belonging to the same flow.

L2 unicast flow learning is very similar to Broadcom L3 unicast flow learning, except that it uses L2-tuple instead of L3/4 tuple for connection lookup and classification.

This is the flow sequence for a new L2 unicast flow:

1st packet:

1. Runner performs a L2 lookup and it will be a flow miss (Runner has not learned the flow yet), and the packet is sent to host.
2. Flow Cache performs a L2 lookup and it will be also a flow miss (Flow Cache has not learned the flow yet), and the packet is sent to Linux bridge.
3. Linux bridge performs L2 lookup, L2 modifies and then forwards the packets to the egress port. When the packet is transmitted, Flow Cache learns the flow.

2nd packet onwards:

1. Runner performs a L2 lookup and it will be a flow hit (Runner has learned this flow). Runner performs L2 modification and then accelerates the packet to the egress port. This packet does not go to Flow Cache and Linux bridge

4 L2 Flow Display

A flow can be an accelerated by flow cache only, or by flow cache and a hardware accelerator (such as the Runner).

4.1 L2 Flow Accelerated by Flow Cache

A user can dump the L2 flows (accelerated by flow cache) on the console using the following command:

```
# cat /proc/fcache/l2list
Broadcom Packet Flow Cache v2.2 May 16 2015 14:17:25
Stats: look<13> fail<23> walk<13> hwm<1> allocs<2> asserts<0>
FlowObject     idle:+swhit SW_TotHits:TotalBytes HW_tpl HW_TotHits  L1-Info      MAC SA
MAC DA          EthType    Vlan0        Vlan1  tag#   IqPrio SkbMark
0xd9c000a0@00001  110:     0          6:       768 0xffff          0  EPHY  8 <00:10:94:00:00:04>
<00:10:94:00:00:01>   0x8847 0xffffffff   0xffffffff   0        1 0x00000000
0xd9c00140@00002  100:     0          0:         0 0xffff          0  EPHY  0 <00:10:94:00:00:01>
<00:10:94:00:00:04>   0x8847 0x81000064   0xffffffff   1        1 0x08000001
```

4.2 L2 Flow Accelerated by Runner

A user can dump the L2 flows (accelerated by Runner) on the console using the following commands:

For DSL platforms:

```
# bs /b/e l2_ucast
```

For PON platforms:

```
# bs /b/e l2_class
```

5 Limitations

These are some of the known limitations for L2 acceleration mode:

- No classification or modification based on L3+L4 fields other than IP-ToS.
- No support for more than two VLAN tag
- Not supported by all hardware accelerators

Revision History

CPE-AN3500, November 15th, 2018

- Initial release