Экстремальный мультикаст

Как это работает

Коммутаторы Extreme для хранения мультикаст таблицы могут использовать как L2-таблицу (FDB), так и L3-таблицу (RIB). Причем ранее, на предыдущих моделях чипов, использовалась только L3-таблица для мультикаст кеша, независимо от того L2 он или L3. На коммутаторах x440 и старше для поиска используется в первую очередь FDB или FDB и RIB.

Коммутация L2 мультикаста: после получения мультикаст пакета, по паре MAC-адрес и номеру VLAN ищется L2 мультикаст таблица, результатом будет индекс на таблицу. Мультикаст таблица возвращает битовую карту исходящих портов для пакета.

L3 мультикаст: После получения мультикаст-пакета, L3 хост таблица ищется по адресу источника, группе и номеру vlan (S,G,V) или по группе, номеру vlan (*, G, V). Соответствующая запись содержит индекс на таблицы мультикаст групп. Результатом L3_IP мультикаст таблицы будет L2 битовая карта портов, в которые пакет должен быть завернут.

Модуль multicast manager имеет кеш и базу данных с индексированными S,G,V. Он взаимодействует с HAL для записи/редактирования/удаления мультикаст содержимого в мультикаст кеша в ASIC.

Алгоритм поиска в кеше может быть изменен на один из следующих:

- source-group-vlan - используется только RIB (L3 multicast lookup) с поиском по юсS,G,V, это режим по умолчанию.
- group-vlan - используется только RIB (L3 multicast lookup) с поиском по *,G,V.
- mac-vlan - используется FDB, поиск по групповым MAC-адресам и номерам VLAN, для L2 мультикаста.
- mixed-mode - используется и FDB и RIB для мультикаста. Логика сохранения в аппаратный кеш такова:
    - При необходимости пересылки мультикаста между VLAN, запись кешируется в RIB, например для PIM, MVR, PVLAN.
    - Любые IPv4/IPv6 адреса, подпадающие под маски MAC-адреса 01:00:5e:00:00:xx, 33:33:00:00:00:xx, 33:33:00:00:01:xx или 33:33:ff:xx:xx:xx, сохраняются в RIB.
    - L2-записи кеша ((IGMP Snooping/PIM snooping/MLD snooping), не подпадающие под маски MAC-адреса 01:00:5e:00:00:xx, 33:33:00:00:00:xx, 33:33:00:00:01:xx или 33:33:ff:xx:xx:xx, сохраняются в FDB.

Любое изменение конфигурации поиска вызывает сброс кеша, и остановку мультикаст-трафика, до момента, пока система вновь не выучит подписки.

Мультикаст-процесс EXOS поддерживает записи кеша "S, G, V", и взаимодействует с HAL в обычном порядке. EXOS hardware abstraction layer (HAL) применяет логику сохранения записей кеша в аппаратные таблицы описанную выше. Если нужно сохранить запись кеша в аппаратную таблицу L2, HAL получает MAC-адрес на основе стандартной логики и устанавливает эту запись MAC в таблице L2.

Для передачи мультикаст-трафика не проверяется отображение (мапинг) мультикаст IP-адреса в MAC-адрес. Если режим поиска настроен как "Mac-VLAN» или «mixed-mode", модуль ядра изменяется для проверки отображения, если пакету не удается пройти через мапинг, то он отбрасывается.

IPMC Compression

Для того чтобы увеличить емкость мультикаст таблицы, в EXOS реализован функционал IPMC Compression, который позволяет сохранять, когда исходящие порты одинаковы, несколько <S,G,V> (or <*,G,V>) IP мультикаст FDB записей в одну и ту же таблицу адреса мультикаст группы. При повторном получении L2 мультикаст записи, компрессия используется заново. В этом случае, несколько записей <MAC,VLAN> многоадресной FDB можно использовать на один индекс L2MC, если исходящие порты записей кэша одинаковы.

Что с этим делать?

Одну из проблем, чаще других попадающуюся, можно описать одной строкой из логов:

<Warn:Kern.IPv4Mc.Warning> IPv4 multicast entry not added. Hardware L3 Table full.

Зная выше описанные принципы работы можно решить эту проблему и многие другие несколькими способами:

1) Мы можем изменить режим работы мультикаст-кеша с source-group-vlan на mac-vlan (или перевести в mixed-mode). Этим мы изменим место хранения кеша с L3 table на FDB. При наличии свободного места в ней, это решит проблему, однако станет невозможным использовать на коммутаторе мультикаст-маршрутизацию.

Для коммутатора x670 данное решение выглядит достаточно привлекательным, т.к. он обладает в значительной степени большей L2 table, чем L3 table.

Начиная с ExOS 15.3:

configure forwarding ipmc lookup-key [group-vlan | source-group-vlan | mac-vlan |

mixed-mode]

До ExOS 15.3:

configure igmp snooping forwarding-mode [group-vlan | source-group-vlan]

Посмотреть состояние можно командой:

show forwarding configuration

2) L3 table хранится в LPM (Longest Prefix Match) таблице коммутатора и делит ее с RIB. По умолчанию большая часть LPM в коммутаторах зарезервирована под RIB. Объем резерва можно посмотреть командой:

show iproute reserved-entries

Например, фактически у нас маршрутов в пределах 700, снизим ограничение до 1000, тем самым освободив память для хранения мультикаст групп:

configure iproute reserved-entries 1000

3) Записи в памяти хешируются, по умолчанию в crc32, можно изменить его на crc16, что так же даст большую вместимость, но повысится процент потерь.

configure forwarding hash-algorithm crc16

Теперь вы знаете, что делать, если у вас появилась такая запись в логе коммутатора:

X670# <Warn:Kern.IPv4Mc.Warning> IPv4 multicast entry not added. Hardware L3 Table full.

Замечание

Опция igmp router-alert.

По умолчанию ExOS принимает и обрабатывает igmp пакеты в соответствии с опцией router-alert в пакете. Так должны работать все нормальные коммутаторы.

По стандарту IETF требуется принятие и обработка IGMPv2 и IGMPv3 пакетов только тогда когда установлена эта опция. Поэтому если в сети установлен D-link или другие коммутаторы вендоров со своим видением стандартов, лучше включить:

configure igmp router-alert receive-required on
configure igmp router-alert transmit on