Internet Engineering Task Force (IETF)
Request for Comments: 7130
Category: Standards Track
ISSN: 2070-1721

M. Bhatia, Ed.
Alcatel-Lucent
M. Chen, Ed.
Huawei Technologies
S. Boutros, Ed.
M. Binderberger, Ed.
Cisco Systems
J. Haas, Ed.
Juniper Networks

February 2014

Двухстороннее детектирование пересылки (BFD) на интерфейсах LAG

Bidirectional Forwarding Detection (BFD) on Link Aggregation Group (LAG) Interfaces

Аннотация

Этот документ определяет механизм для запуска двухстороннего детектирования пересылки (Bidirectional Forwarding Detection или BFD) на интерфейсах LAG[1]. Это выполняется путём запуска независимой асинхронной сессии BFD на каждом канале в составе LAG.

Механизм позволяет проверить целостность канала при наличии или отсутствии протокола LACP[2]. Механизм обеспечивает более короткое время детектирования по сравнению с LACP. Проверка целостности охватывает также элементы двухсторонней пересылки сетевого уровня (L3).

Статус документа

Документ относится к категории Internet Standards Track.

Документ является результатом работы IETF[3] и представляет согласованный взгляд сообщества IETF. Документ прошёл открытое обсуждение и был одобрен для публикации IESG[4]. Дополнительную информацию о стандартах Internet можно найти в разделе 2 RFC 5741.

Информацию о текущем статусе документа, ошибках и способах обратной связи можно найти по ссылке http://www.rfc-editor.org/info/rfc7130.

Оглавление

1. Введение

Протокол BFD [RFC5880] обеспечивает механизм потенциально быстрого обнаружения отказов на двухсторонних путях между парой машин пересылки, включая интерфейсы и каналы передачи данных, а также, насколько это возможно, сами машины. Протокол BFD также обеспечивает быстрый механизм обнаружения коммуникационных отказов на любых каналах данных и может работать через любую среду на любом физическом уровне.

[1]Link Aggregation Group - группа агрегированных каналов.
[2]Link Aggregation Control Protocol - протокол управления объединением каналов.
[3]Internet Engineering Task Force - комиссия по решению инженерных задач Internet.
[4]Internet Engineering Steering Group - комиссия по инженерным разработкам Internet.

LAG в соответствии с определением [IEEE802.1AX] обеспечивает механизмы объединения множества физических каналов в один логический. Этот логический канал обеспечивает большую пропускную способность и надёжность, поскольку при отказе одного из физических каналов объединённый логический канал может продолжать пересылку трафика через оставшиеся физические каналы.

В настоящее время для обнаружения отказов на физических каналах группы используется протокол LACP. Однако использование BFD обеспечивает (1) более быстрое обнаружение, (2) не требует наличия LACP и (3) позволяет проверить способность каждого канала группы пересылать пакеты L3.

Организация одной сессии BFD на объединённом канале без информации о членах группы не позволяет BFD гарантированно обнаруживать отказы физических каналов данной группы.

Цель состоит в проверке связности (Continuity) для каждого канала в группе. Это соответствует параграфу 7.3 в [RFC5882].

Предложенный в этом документе подход заключается в организации асинхронной сессии BFD на каждом канале группы LAG и BFD-управления включением члена LAG в таблицу балансировки нагрузки L2 на интерфейсе LAG при наличии или отсутствии LACP.

Документ описывает организацию асинхронной сессии BFD на физических каналах группы интерфейса LAG.

Хотя в Ethernet имеются свои механизмы детектирования отказов (802.1ax, .3ah), которые можно использовать для LAG, предлагаемое здесь решение позволяет операторам, которые уже развернули BFD для других технологий (например, IP или MPLS) применять для детектирования отказов общий механизм.

1.1. Уровни требований

Ключевые слова **необходимо** (MUST), **недопустимо** (MUST NOT), **требуется** (REQUIRED), **нужно** (SHALL), **не нужно** (SHALL NOT), **следует** (SHOULD), **не следует** (SHOULD NOT), **рекомендуется** (RECOMMENDED), **возможно** (MAY), **необязательно** (OPTIONAL) в данном документе интерпретируются в соответствии с RFC 2119 [RFC2119].

2. BFD на каналах LAG

Механизм быстрого обнаружения отказов на каналах группы LAG основан на организации асинхронной сессии BFD для каждого физического канала LAG. Такие сессии далее называются сессиями micro-BFD.

2.1. Семейство адресов сессий Micro-BFD

Сессии micro-BFD на каналах групп при использовании инкапсуляции IP/UDP, могут применять адреса IPv4 или IPv6. На канале группы **могут** существовать две сессии micro-BFD - одна IPv4, другая IPv6. При использовании семейства адресом на канале группы, оно **должно** применяться для всех каналов данной группы LAG.

2.2. Согласование сессии Micro-BFD

Для каждого разрешённого семейства адресов организуется одна сессия micro-BFD на каждом канале группы LAG. Согласование сессий micro-BFD **должно** выполняться в соответствии с процедурой, определённой в [RFC5880] и [RFC5881].

В этом документе рассматривается лишь асинхронный режим BFD, а использование эхо-функции BFD выходит за рамки документа. По крайней мере одна из систем **должна** быть активной (Active role). Сессии micro-BFD на каналах группы являются независимыми сессиями BFD. Они используют свои уникальные дискриминаторы, поддерживают свои наборы переменных состояния и имеют свои независимые машины состояний. Значения таймеров **могут** быть разными для для сессий micro-BFD, относящихся к одной группе, хотя предполагаются одинаковые значения таймеров для всех сессий micro-BFD одной группы каналов.

Демультиплексирование полученного пакета BFD основано исключительно на значениях поля Your Discriminator, если оно отлично от 0. Для начальных пакетов Down BFD в сессии BFD это поле **может** иметь нулевое значение. В таких случаях демультиплексирование **должно** выполняться на основе комбинации других полей, которая **должна** включать информацию об интерфейсе канала группы и номер порта получателя UDP из полученного пакета BFD.

Процедура приёма управляющих пакетов BFD, описанная в параграфе 6.8.6 [RFC5880] переопределена для сессий micro-BFD на каналах группы LAG, как указано ниже.

> Если поле Your Discriminator отлично от 0 и обнаружена сессия micro-BFD для канала LAG, интерфейс, на котором принят пакет управления micro-BFD, **должен** соответствовать связанному с этой сессией интерфейсу.

Этот документ определяет, что пакеты управления BFD для каждой сессии micro-BFD инкапсулируются в IP/UDP, как определено в [RFC5881], но используется порт получателя UDP 6784.

Использование нового порта UDP позволяет отличать BFD через LAG от BFD через один интервал (single-hop) IP. Примером такой ситуации может служить (некорректно) настроенная группа LAG с сессиями micro-BFD на одной стороне, использующая сессию [RFC5881] BFD для LAG (считается одним интерфейсом) на другой стороне.

Описанные в этом документе процедуры **должны** применяться для сообщений BFD, адресованных в порт 6784, и их **недопустимо** использовать для других портов, указанных в RFC для других режимов BFD.

Пакеты управления используют IP-адрес получателя, который настроен на партнерской системе и доступен через интерфейс LAG.

Реализации могут варьироваться от явной настройки адресов IP для сессий BFD до использования специальных (out-of-band) методов определения IP-адреса получателя. Детали этого выходят за рамки документа.

2.3. Сессии Micro-BFD на каналах Ethernet

На каналах Ethernet группы LAG в качестве MAC-адреса получателя используется выделенный групповой адрес MAC 01-00-5E-90-00-01 для ближайшего интервала пересылки. Выделенный MAC-адрес **должен** применяться для первоначальных пакетов BFD сессии micro-BFD в состояниях Down/AdminDown и Init. При переходе сессии micro-BFD в состояние Up, первые пакеты bfd.DetectMult для состояния Up **должны** передаваться по выделенному адресу MAC. Для пакетов BFD в состоянии Up, следующих за первыми пакетами bfd.DetectMult, **может** указываться MAC-адрес отправителя из полученного пакета BFD для этой сессии вместо использования выделенного MAC-адреса.

Все реализации **должны** быть способны принимать и передавать пакеты BFD в состоянии Up с использованием выделенного MAC-адреса. Реализациям, поддерживающим передачу пакетов BFD Up с выделенным и полученным MAC-адресом, следует обеспечить способ выбора поведения.

На Ethernet-каналах группы LAG в качестве MAC-адреса отправителя **следует** использовать MAC-адрес физического интерфейса, передающего пакет.

Этот механизм позволяет снизить потребность в дополнительных MAC-адресах, что ведёт к снижению потребности в ресурсах Ethernet на приёмной стороне входящих в группу каналов.

Пакеты сессий Micro-BFD **следует** во всех случаях передавать без тегов. Однако при использовании LAG в контексте IEEE 802.1q или IEEE 802.qinq пакеты micro-BFD могут передаваться без тегов или с нулевым значением тега vlan (тег приоритета 802.1p). Реализации, соответствующие данному стандарту, **должны** быть способны принимать оба типа пакетов micro-BFD.

3. Взаимодействие LAG и BFD

Сессии micro-BFD для определённого члена LAG **должны** запрашиваться, когда этот канал находится в состоянии Distributing или Standby. Сессии **должны** удаляться, когда состояние канала отличается от Distributing и Standby.

BFD используется для управления, если алгоритм распределения нагрузки способен выбирать отдельный канал в LAG. Иными словами, даже при использовании протокола LACP и готовности членов группы пересылать трафик балансировщику нагрузки **недопустимо** использовать канал, пока все сессии micro-BFD на этом канале не находятся в состоянии Up.

Если реализация имеет отдельные таблицы балансировки для IPv4 и IPv6 и на канале имеются сессии micro-BFD для IPv4 и IPv6, реализация **может** разрешить использование канала механизмом распределения нагрузки на основе сессии BFD для соответствующего семейства адресов.

Исключением являются сами пакеты BFD. Реализации **могут** принимать и передавать пакеты BFD через интерфейс MAC-сервиса агрегатора независимо от состояния сессии.

4. BFD на каналах LAG и приложения L3

Описанный в этом документе механизм скорей всего будет применяться модулями, управляющими интерфейсами или группами LAG, управляя тем самым отдельными каналами LAG. Типовые протоколы L3 (например, OSPF) не имеют представления о LAG и рассматривают его как один «большой» интерфейс. Сигнализация от микросессий в протоколы L3 реально осуществляется с помощью сессий micro-BFD в таблице балансировки нагрузки и возможного решения модуля управления Interface/LAG о закрытии (shut down) LAG. Активный метод проверки влияния сессий micro-BFD для протоколов L3 состоит в запросе одной сессии BFD на LAG.

5. Детектирование отказа члена группы

При отключении сессии micro-BFD соответствующий канал группы **должен** удаляться из таблиц(ы) распределения нагрузки LAG.

Если реализация использует разные таблицы балансировки нагрузки для IPv4 и IPv6 и на канале существуют сессии micro-BFD для IPv4 и IPv6, реализация **может** удалять канал лишь из той таблицы распределения нагрузки для того семейства адресов, сессия которого была разорвана. Например, сессия IPv4 micro-BFD может отказать, а сессия IPv6 micro-BFD остаться в состоянии Up и канал **может** быть удалён из таблицы балансировки IPv4, но **может** быть сохранен в таблице балансировки IPv6. Допускается и удаление в таких случаях канала из обеих таблиц. Выбор решения определяется реализацией.

6. Вопросы безопасности

Этот не добавляет проблем безопасности, а механизмы защиты, описанные в [RFC5880], применимы к нему.

7. Взаимодействие с IANA

Агентство IANA назначило выделенный MAC-адрес 01-00-5E-90-00-01 (см. [RFC7042]), а также порт UDP 6784 для двухстороннего детектирования пересылки (BFD) на интерфейсах LAG. Была также указана ссылка на [RFC7130].

Агентство IANA изменило запись в реестре для порта 6784, чтобы показать выделение номера [IESG] и указать в качестве контакта [BFD_Chairs]. Преобразование [BFD_Chairs] указано как mailto:bfd-chairs@tools.ietf.org. Указана также ссылка на [RFC7130].

8. Благодарности

Авторы благодарны Dave Katz, Alexander Vainshtein, Greg Mirsky и Jeff Tantsura за их отзывы.

Началом текущего обсуждения послужило распространение документа Bidirectional Forwarding Detection (BFD) for Interface (июль 2011 г.).

9. Участники работы

Paul Hitchen
BT
EMail: paul.hitchen@bt.com

George Swallow
Cisco Systems
EMail: swallow@cisco.com

Wim Henderickx
Alcatel-Lucent
EMail: wim.henderickx@alcatel-lucent.com

Nobo Akiya
Cisco Systems
EMail: nobo@cisco.com

Neil Ketley
Cisco Systems
EMail: nketley@cisco.com

Carlos Pignataro
Cisco Systems
EMail: cpignata@cisco.com

Nitin Bahadur
Bracket Computing
EMail: nitin@brkt.com

Zuliang Wang
Huawei Technologies
EMail: liang_tsing@huawei.com

Liang Guo
China Telecom
EMail: guoliang@gsta.com

Jeff Tantsura
Ericsson
EMail: jeff.tantsura@ericsson.com

10. Литература

10.1. Нормативные документы

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, June 2010.

[RFC5881] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD) for IPv4 and IPv6 (Single Hop)", RFC 5881, June 2010.

[RFC5882] Katz, D. and D. Ward, "Generic Application of Bidirectional Forwarding Detection (BFD)", RFC 5882, June 2010.

10.2. Дополнительная литература

[IEEE802.1AX] IEEE Std. 802.1AX, "IEEE Standard for Local and metropolitan area networks - Link Aggregation", November 2008.

[RFC7042] Eastlake, D. and J. Abley, "IANA Considerations and IETF Protocol and Documentation Usage for IEEE 802 Parameters", BCP 141, RFC 7042, October 2013.

Приложение A. Вопросы использования BFD на каналах группы

Если функция BFD-over-LAG начала предоставляться на канале группы уже после активизации канала в LAG, состояние сессии BFD не следует учитывать в алгоритме распределения нагрузки, пока сессия BFD не перейдёт в состояние Up. Если сессия BFD не перешла в состояние Up, но группа LAG стала неактивной, можно применять описанные ранее процедуры.

Это обеспечивает отсутствие влияния последовательности событий «включение LAG, затем включение BFD на LAG» на службы пересылки.

Если функция BFD-over-LAG перестала предоставляться на канале группы, когда связанная с ним сессия micro-BFD находилась в состоянии Up, BFD следует перейти в состояние AdminDown и предпринять попытку передать это состояние партнёру.

Если локальным или удаленным состоянием сессии micro-BFD является AdminDown, системе не следует показывать отказ соединения какому-либо клиенту и не следует удалять конкретный канал группы LAG из процесса пересылки. Это поведение не зависит от использования протокола LACP для LAG.

Когда трафик пересылается через канал в то время, как соответствующая сессия micro-BFD не находится в состоянии Up, реализация может использовать настраиваемый тайм-аут после которого сессия BFD должна перейти в состояние Up или пересылка через канал должна прекратиться.

При наличии такого тайм-аута должны обеспечиваться функция, позволяющая включать и отключать его применение.

Значение настраиваемого тайм-аута должно гарантировать, что LAG не останется навсегда в «несогласованном» состоянии, когда пересылка через канал происходит без подтверждения сессии micro-BFD о нормальной работе канала.

Когда устройство не использует сессию micro-BFD на канале в то время, как другое устройство делает это и воспринимает сессию как отключённую (Down), это будет приводить к двум разным представлениям о состоянии канала. Скорей всего это приведёт к потере трафика в LAG. Использование другого протокола для загрузки BFD может обнаружить такое рассогласование конфигурации, поскольку не настроенная сторона будет возвращать ошибку (reject). Такие механизмы загрузки выходят за рамки этого документа.

Адреса авторов

Manav Bhatia (редактор)
Alcatel-Lucent
Bangalore 560045
India
EMail: manav.bhatia@alcatel-lucent.com

Mach(Guoyi) Chen (редактор)
Huawei Technologies
Q14 Huawei Campus, No. 156 Beiqing Road, Hai-dian District
Beijing 100095
China
EMail: mach@huawei.com

Sami Boutros (редактор)
Cisco Systems
EMail: sboutros@cisco.com

Marc Binderberger (редактор)
Cisco Systems
EMail: mbinderb@cisco.com

Jeffrey Haas (редактор)
Juniper Networks
EMail: jhaas@juniper.net

Перевод на русский язык

Николай Малых
nmalykh@protokols.ru