

Internet Engineering Task Force (IETF)  
Request for Comments: 8971  
Category: Informational  
ISSN: 2070-1721

S. Pallagatti, Ed.  
VMware  
G. Mirsky, Ed.  
ZTE Corp.  
S. Paragiri  
Individual Contributor  
V. Govindan  
M. Mudigonda  
Cisco  
December 2020

## Bidirectional Forwarding Detection (BFD) for Virtual eXtensible Local Area Network (VXLAN)

Обнаружение двухсторонней пересылки для VXLAN

### Аннотация

Этот документ описывает применение протокола обнаружения двухсторонней пересылки (BFD<sup>1</sup>) в туннелях «точка-точка» расширяемых виртуальных ЛВС (VXLAN<sup>2</sup>), используемых для формирования наложенной сети.

### Статус документа

Документ не относится к категории Internet Standards Track и публикуется лишь для информации.

Документ является результатом работы IETF<sup>3</sup> и представляет согласованный взгляд сообщества IETF. Документ прошёл открытое обсуждение и был одобрен для публикации IESG<sup>4</sup>. Дополнительную информацию о стандартах Internet можно найти в разделе 2 RFC 7841.

Информацию о текущем статусе документа, ошибках и способах обратной связи можно найти по ссылке <https://www.rfc-editor.org/info/rfc8971>.

### Авторские права

Авторские права (Copyright (c) 2020) принадлежат IETF Trust и лицам, указанным в качестве авторов документа. Все права защищены.

К документу применимы права и ограничения, указанные в BCP 78 и IETF Trust Legal Provisions и относящиеся к документам IETF (<http://trustee.ietf.org/license-info>), на момент публикации данного документа. Прочтите упомянутые документы внимательно. Фрагменты программного кода, включённые в этот документ, распространяются в соответствии с упрощённой лицензией BSD, как указано в параграфе 4.e документа IETF Trust Legal Provisions, без каких-либо гарантий (как указано в Simplified BSD License).

## Оглавление

|   |   |
|---|---|
| 1. Введение.....                                | 1 |
| 2. Используемые соглашения.....                 | 2 |
| 2.1. Сокращения.....                            | 2 |
| 2.2. Уровни требований.....                     | 2 |
| 3. Развёртывание.....                           | 2 |
| 4. Использование Management VNI.....            | 3 |
| 5. Передача пакета BFD через туннель VXLAN..... | 3 |
| 6. Приём пакета BFD из туннеля VXLAN.....       | 4 |
| 7. Echo BFD.....                                | 4 |
| 8. Взаимодействие с IANA.....                   | 4 |
| 9. Вопросы безопасности.....                    | 4 |
| 10. Литература.....                             | 4 |
| 10.1. Нормативные документы.....                | 4 |
| 10.2. Дополнительная литература.....            | 4 |
| Благодарности.....                              | 5 |
| Участник работы.....                            | 5 |
| Адреса авторов.....                             | 5 |

## 1. Введение

В документе «Virtual eXtensible Local Area Network (VXLAN)» [RFC7348] представлена схема инкапсуляции, позволяющая создавать наложенную сеть за счёт отвязывания адресного пространства подключённых хостов от сети.

Одним из применений VXLAN является соединение виртуальных машин арендатора (VM) в центре обработки данных (ЦОД). VXLAN отвечает требованиям к сетевой инфраструктуре ЦОД L2 и L3 при работе VM в среде с разными арендаторами за счёт схемы наложения уровня L2 на сеть L3 [RFC7348]. Другим применением является инкапсуляция Ethernet VPN [RFC8365].

<sup>1</sup>Bidirectional Forwarding Detection.

<sup>2</sup>Virtual eXtensible Local Area Network.

<sup>3</sup>Internet Engineering Task Force - комиссия по решению инженерных задач Internet.

<sup>4</sup>Internet Engineering Steering Group - комиссия по инженерным разработкам Internet.

В документе предполагается использование VXLAN для виртуализованных хостов и речь идёт о VM и конечных точках туннелей VXLAN (VXLAN Tunnel End Points или VTEP) в гипервизорах. Однако эти же концепции применимы к не виртуализованным хостам, подключённым к точкам VTEP в коммутаторах.

При отсутствии маршрутизаторов в наложенной сети VM может взаимодействовать с другими VM лишь из того же сегмента VXLAN. VM не знают о туннелях VXLAN, поскольку эти туннели завершаются (terminate) в точках VTEP, которые обеспечивают инкапсуляцию и декапсуляцию кадров, передаваемых между VM.

Способность отслеживать неразрывность пути, т. е. упреждающая проверка непрерывности (continuity check или CC) для туннелей VXLAN «точка-точка» (p2p) имеет важное значение. Для мониторинга туннелей p2p VXLAN применяется асинхронный режим BFD, заданный в [RFC5880].

При участии в VXLAN узлов группового сервиса (Multicast Service Node или MSN), как описано в параграфе 3.3 [RFC8293], применяется описанный здесь механизм, который может служить для проверки непрерывности пути между исходной точкой виртуализации сети (Network Virtualization Endpoint или NVE) и MSN.

Этот документ описывает использование протокола BFD для мониторинга неразрывности пути между точками VXLAN VTEP, которые служат VNE, и/или между NVE-источником и репликатором MSN с использованием идентификатора сети VXLAN (VXLAN Network Identifier или VNI), описанного в разделе 4. Проверки для других конечных точек VXLAN выходят за рамки спецификации.

## 2. Используемые соглашения

### 2.1. Сокращения

**BFD** *Bidirectional Forwarding Detection*

Обнаружение двухсторонней пересылки.

**CC** *Continuity Check*

Проверка неразрывности.

**FCS** *Frame Check Sequence*

Последовательность проверки кадра.

**MSN** *Multicast Service Node*

Узел группового сервиса (службы).

**NVE** *Network Virtualization Endpoint*

Конечная точка виртуализации сети.

**p2p** *Point-to-point*

Соединение «точка-точка».

**VFI** *Virtual Forwarding Instance*

Виртуальный экземпляр пересылки.

**VM** *Virtual Machine*

Виртуальная машина.

**VNI** *VXLAN Network Identifier (VXLAN Segment ID)*

Идентификатор сети (сегмента) VXLAN.

**VTEP** *VXLAN Tunnel End Point*

Конечная точка туннеля VXLAN.

**VXLAN** *Virtual eXtensible Local Area Network*

Расширяемая виртуальная ЛВС.

### 2.2. Уровни требований

Ключевые слова **необходимо** (MUST), **недопустимо** (MUST NOT), **требуется** (REQUIRED), **нужно** (SHALL), **не следует** (SHALL NOT), **следует** (SHOULD), **не нужно** (SHOULD NOT), **рекомендуется** (RECOMMENDED), **не рекомендуется** (NOT RECOMMENDED), **возможно** (MAY), **необязательно** (OPTIONAL) в данном документе интерпретируются в соответствии с BCP 14 [RFC2119] [RFC8174] тогда и только тогда, когда они выделены шрифтом, как показано здесь.

## 3. Развёртывание

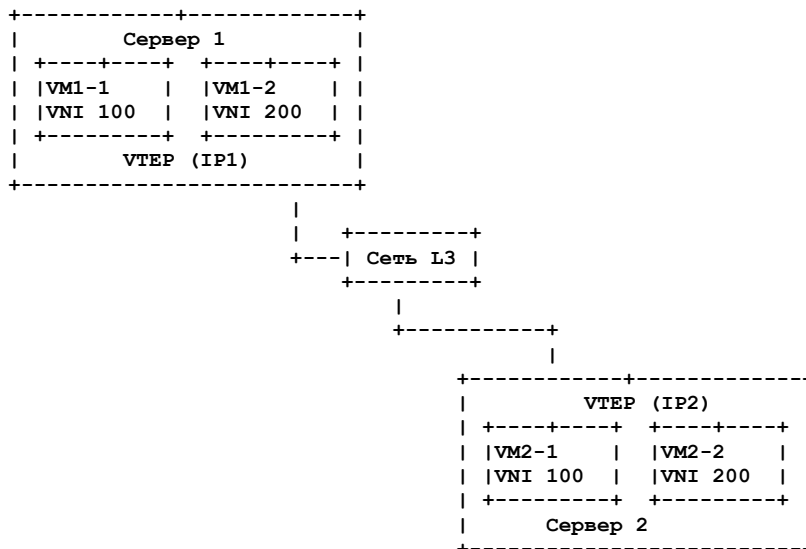


Рисунок 1. Домен VXLAN.

На рисунке 1 показан случай с 2 серверами, на каждом из которых поддерживается по 2 VM. Серверы поддерживают точки VTEP, завершающие туннели VXLAN, со значениями VNI 100 и 200. Между VTEP могут быть организованы отдельные сессии BFD (IP1 и IP2) для мониторинга каждого из туннелей VXLAN (VNI 100 и VNI 200). Применение сессии BFD для мониторинга набора VXLAN VNI между парой точек VTEP может помочь при обнаружении и локализации проблем, связанных с ошибками в конфигурации. Реализация, поддерживающая эту спецификацию, **должна** быть способна контролировать число сессий BFD, которые могут быть созданы между парой точек VTEP. Метод применим для виртуальных и физических точек VTEP.

Одновременно может использоваться сессия BFD сервисного уровня между точками VTEP арендаторов (IP1 и IP2) для сквозного контроля отказов, но этот вопрос выходит за рамки документа. В этом случае для точек VTEP пакеты BFD Control данной сессии не отличимы от пакетов данных.

Для пакетов BFD Control, инкапсулированных в VXLAN (рисунок 2), для внутреннего IP-адреса получателя **следует** указывать один из loopback-адресов (127/8 для IPv4) или один из отображаемых на IPv4 loopback-адресов IPv6 (::ffff:127.0.0.0/104 для IPv6).

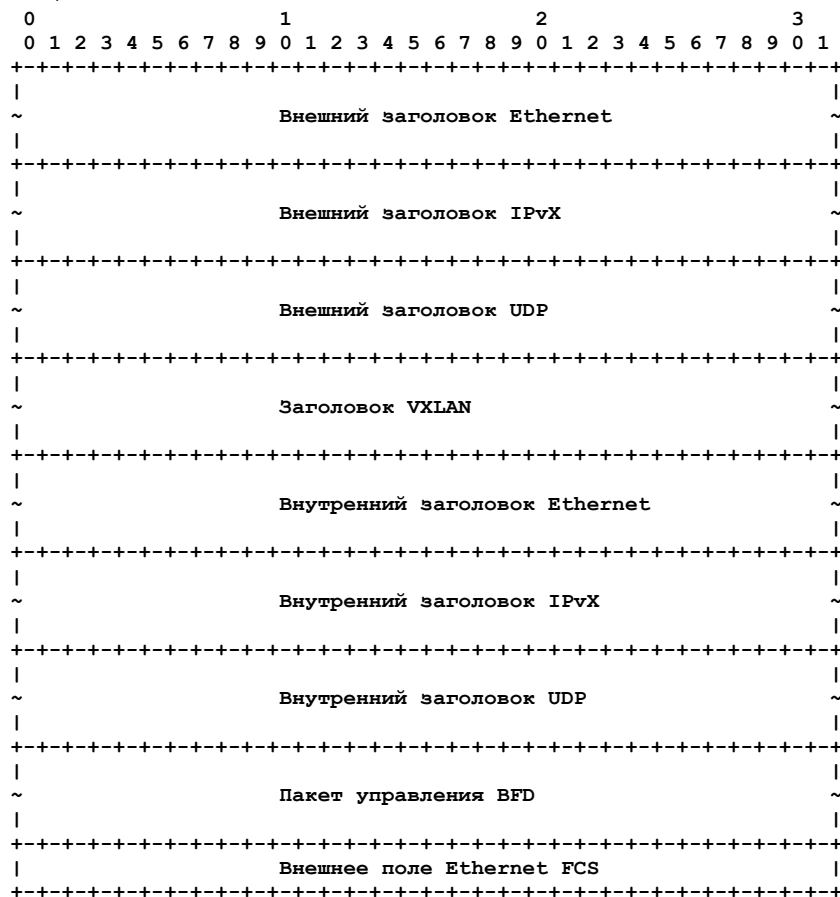


Рисунок 2. VXLAN-инкапсуляция пакета управления BFD.

#### 4. Использование Management VNI

В большинстве случаев данной точке VTEP достаточно одной сессии BFD для мониторинга доступности удалённой VTEP, независимо от числа VNI. Управляющие сообщения BFD **должны** передаваться с использованием Management VNI, служащего каналом управления и поддержки между VTEP. Реализация **может** поддерживать работу BFD с другим (не Management) VNI, но этот вопрос выходит за рамки документа. Выбор номера VNI для Management VNI **должен** задаваться плоскостью управления. Реализация **может** использовать VNI 1 в качестве принятого по умолчанию номера Management VNI. Все пакеты VXLAN, принятые с Management VNI, **должны** обрабатываться локально и **недопустимо** пересылать их арендатору.

#### 5. Передача пакета BFD через туннель VXLAN

Пакеты BFD **должны** инкапсулироваться и передаваться удалённой точке VTEP, как описано в этом параграфе. Реализации **следует** обеспечивать следование пакетов BFD по тому же пути пересылки, который применяется для пакетов данных VXLAN внутри системы отправителя.

Инкапсуляция пакетов BFD в VXLAN описана ниже. Формат пакетов VXLAN определён в разделе 5 [RFC7348]. Поле VNI в заголовке VXLAN **должно** иметь значение, выбранное для Management VNI. Внешние заголовки IP/UDP и заголовок VXLAN **должны** устанавливаться отправителем в соответствии с [RFC7348].

Пакет BFD **должен** передаваться во внутреннем кадре Ethernet пакета VXLAN. Выбор адресов получателя (MAC и IP) для внутреннего кадра Ethernet **должен** гарантировать, что пакет BFD Control не будет пересылаться арендатору, но будет обработан локально удалённой точкой VTEP. Поля внутреннего кадра Ethernet с пакетом BFD Control описаны ниже.

##### Заголовок Ethernet Destination MAC

Значение Management VNI, которое не использует ни один арендатор, не будет иметь выделенного MAC-адреса для декапсулированного трафика. В этом поле **следует** устанавливать значение 00-52-02.

**Source MAC**

MAC-адрес, связанный с отправляющей точкой VTEP.

**Ethertype**

Устанавливается 0x0800 для внутреннего заголовка IPv4 или 0x86DD для IPv6.

**Заголовок IP****Destination IP**

В этом поле **недопустимо** оказывать какой-либо из IP-адресов арендатора. Адрес IP **следует** брать из диапазона 127/8 для IPv4 или из диапазона ::ffff:127.0.0.0/104 для IPv6. Кроме того, это поле **может** содержать IP-адрес точки VTEP.

**Source IP**

IP-адрес отправляющей точки VTEP.

**TTL или Hop Limit**

В соответствии с [RFC5881] поле **должно** иметь значение 255.

Для порта получателя UDP устанавливается значение 3784, а поля пакета BFD Control устанавливаются в соответствии с [RFC5881].

## 6. Приём пакета BFD из туннеля VXLAN

При получении пакета точка VTEP **должна** проверить его. Если пакет получен с Management VNI и определён как пакет BFD Control, адресованный VTEP, выполняется обработка пакета. Обработка пакетов BFD Control, принятых не с Management VNI, выходит за рамки документа.

Затем проверяется содержимое внутренних данных (payload) пакета IP в соответствии с разделами 4 и 5 в [RFC5881].

## 7. Echo BFD

Поддержка Echo BFD выходит за рамки этого документа.

## 8. Взаимодействие с IANA

Агентство IANA выделило MAC-адрес со значением 00-52-02 из блока «Unassigned (small allocations)» в реестре «IANA Unicast 48-bit MAC Addresses» с указанием в поле Usage значения «BFD for VXLAN» и ссылкой на этот документ в поле Reference.

## 9. Вопросы безопасности

К этому документу применимы вопросы безопасности, рассмотренные в [RFC5880], [RFC5881] и [RFC7348].

Документ рекомендует применять адреса из блока 127/8 для IPv4 и отображаемые на IP4 loopback-адреса из ::ffff:127.0.0.0/104 для IPv6 в качестве адреса получателя по внутреннему заголовку IP. Использование таких адресов предотвращает пересылку инкапсулированных управляющих сообщений BFD промежуточными узлами в случае разрыва туннеля VXLAN, как указано в [RFC1812].

Маршрутизатору **не следует** пересылать любые пакеты, исходящие из сети 127 (за исключением пакетов, передаваемых через интерфейс loopback). Маршрутизатор **может** поддерживать параметр, который позволяет администратору отключать такие проверки, но по умолчанию эти проверки **должны** выполняться.

Использование отображаемых на IPv4 адресов IPv6 обеспечивает такие же свойства, как использование для IPv4 адресов 127/8. Кроме того, префикс отображаемых на IPv4 адресов IPv6 не анонсируют протоколы маршрутизации.

Если реализация поддерживает создание нескольких сессий BFD между парой точек VTEP, ей **следует** обеспечивать механизм контроля числа активных одновременно сессий.

## 10. Литература

### 10.1. Нормативные документы

[RFC1812] Baker, F., Ed., "Requirements for IP Version 4 Routers", [RFC 1812](#), DOI 10.17487/RFC1812, June 1995, <<https://www.rfc-editor.org/info/rfc1812>>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, [RFC 2119](#), DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.

[RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", [RFC 5880](#), DOI 10.17487/RFC5880, June 2010, <<https://www.rfc-editor.org/info/rfc5880>>.

[RFC5881] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD) for IPv4 and IPv6 (Single Hop)", [RFC 5881](#), DOI 10.17487/RFC5881, June 2010, <<https://www.rfc-editor.org/info/rfc5881>>.

[RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", [RFC 7348](#), DOI 10.17487/RFC7348, August 2014, <<https://www.rfc-editor.org/info/rfc7348>>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, [RFC 8174](#), DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.

### 10.2. Дополнительная литература

[RFC8293] Ghanwani, A., Dunbar, L., McBride, M., Bannai, V., and R. Krishnan, "A Framework for Multicast in Network Virtualization over Layer 3", RFC 8293, DOI 10.17487/RFC8293, January 2018, <<https://www.rfc-editor.org/info/rfc8293>>.

[RFC8365] Sajassi, A., Ed., Drake, J., Ed., Bitar, N., Shekhar, R., Uttaro, J., and W. Henderickx, "A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)", RFC 8365, DOI 10.17487/RFC8365, March 2018, <<https://www.rfc-editor.org/info/rfc8365>>.

## **Благодарности**

Авторы благодарны Jeff Haas из Juniper Networks за рецензии и отклики.

Спасибо Nobo Akiya, Marc Binderberger, Shahram Davari, Donald E. Eastlake 3rd, Anoop Ghanwani, Dinesh Dutt, Joel Halpern, Carlos Pignataro за обширные обзоры, а также подробные и конструктивные комментарии.

## **Участник работы**

**Reshad Rahman**

Cisco

Email: [rrahman@cisco.com](mailto:rrahman@cisco.com)

## **Адреса авторов**

**Santosh Pallagatti (редактор)**

VMware

Email: [santosh.pallagatti@gmail.com](mailto:santosh.pallagatti@gmail.com)

**Greg Mirsky (редактор)**

ZTE Corp.

Email: [gregimirsky@gmail.com](mailto:gregimirsky@gmail.com)

**Sudarsan Paragiri**

Individual Contributor

Email: [sudarsan.225@gmail.com](mailto:sudarsan.225@gmail.com)

**Vengada Prasad Govindan**

Cisco

Email: [venggovi@cisco.com](mailto:venggovi@cisco.com)

**Mallik Mudigonda**

Cisco

Email: [mmudigon@cisco.com](mailto:mmudigon@cisco.com)

## **Перевод на русский язык**

Николай Малых

[nmalykh@protokols.ru](mailto:nmalykh@protokols.ru)