ИНФОРМАТИКА

УДК 004.7

ИНКАПСУЛЯЦИЯ МАГИСТРАЛЬНОГО ТРАФИКА ЦЕНТРА ОБРАБОТКИ ДАННЫХ

А.В. Воруев, О.М. Демиденко, В.Д. Левчук, П.Л. Чечет

Гомельский государственный университет им. Ф. Скорины

ENCAPSULATION OF BACKBONE TRAFFIC OF DATA PROCESSING CENTER

A.V. Varuyeu, O.M. Demidenko, V.D. Liauchuk, P.L. Chechat

F. Scorina Gomel State University

Рассматривается процесс создания виртуальной магистральной линии на базе протоколов оверлейной связи в условиях трансляции трафика через публичные каналы передачи данных.

***Ключевые слова**: топология, медиа-контент, магистральный трафик, VXLAN, VTEP, EVE-NG, канал передачи данных.*

The process of creating a virtual backbone line based on protocols of an overlay communication is considered. The traffic transference is limited to public data channels.

***Keywords**: topology, media content, backbone traffic, VXLAN, VTEP, EVE-NG, channel of data transference.*

Введение

Увеличение количества передаваемых данных в сети связано с развитием IoT среды, когда каждое устройство имеет доступ к сети, например, домашняя онлайн видео-камера, либо домашнее охранное устройство. Большая часть трафика приходится на развлекательный медиаконтент, а именно – на видео. Помимо YouTube, происходит добавление медиа-содержимого в крупнейшие социальные сети, новостные каналы начинают транслировать передачи в режиме реального времени.

Причина этого кроется также в дальнейшем развитии медиа-кодеков и увеличении экранов смартфонов и планшетов, когда есть возможность смотреть видео в 2k/4k-формате. Прогнозируется увеличение доли видео-данных с 70 до 82 процентов, в то время как объём увеличится в 4 раза. За последний год объем трафика видеонаблюдения практически удвоился, а к 2020 г. вырастет десятикратно. Ожидается интенсивное развитие рынка дополненной и виртуальной реальности. За последний год трафик этого вида увеличился в 4 раза и прогнозируется, что к 2020 году он возрастет в 61 раз.

Наметилась тенденция к высокомасштабному увеличению общего объёма передаваемых данных, среди которого будет присутствовать трафик, чувствительный к задержкам. [1], [2] В этом случае рекомендуется разбивать поток данных на несколько меньших и вводить его в облако провайдера на ближайших к абоненту точках. Это необходимо для выполнения балансировки и снижения нагрузки на единый центр обработки данных. В этом случае возможно реализовывать различные сценарии как работы, так и защиты от увеличивающихся размеров DDoS-аттак.

Если два центра обработки данных находятся в одном здании, но в разных его частях, то можно использовать классические методы передачи данных – 1/10/40/100 Гбит Ethernet для обычного сетевого трафика и 8/16/32/128 Гбит Fibre-Channel для сетей хранения данных. Для более сложных случаев можно использовать конвергентные способы передачи данных, т. е. когда транспортная сеть обслуживает одновременно и обычный трафик, и трафик систем хранения данных.

Последовательность шагов по оптимизации сетевой архитектуры может быть следующей:

1. Обновление кабельной инфраструктуры (уменьшение средней длины патч-корда, уменьшение объёма аплинков через использование уплотнительной аппаратуры, либо через использование конвергентных решений).
2. Внедрение оборудования, позволяющего создавать программно-определяемые решения с использованием универсального оборудования.
3. Использование новейших операторских решений по оптимальному использованию оверлейных сетей и протокола BGP.

1 Перспектива развития оверлейных сетей

В данный момент многие архитекторы озабочены созданием нового универсального протокола, который бы объединял все плюсы существующих оверлейных сетей. Название данного проекта – протокол GENEVE (Generic Network Virtualization Encapsulation). Разработчики спецификаций этого протокола стремятся к унификации топологии (рисунок 1.1) [3].

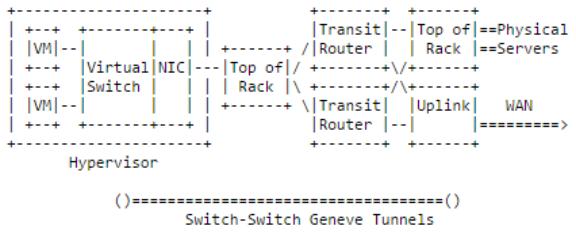

Рисунок 1.1 – Упрощённая схема взаимодействия частей ЦОД

В данный момент самым эффективным механизмом применения всех опций протокола является TLV-формат (type-length-value).

В качестве базового механизма передачи служебных данных GENEVE выбран UDP. Организация IANA также зарезервировала UDP-порт 6081 для работы GENEVE. В связи с тем, что GENEVE есть наследник VXLAN/NVGRE/STT, его интеграция в любое из существующих решений должно пройти достаточно гладко и быстро, т. к. гипервизоры смогут работать одновременно с несколькими протоколами – старым и новым, без какого-либо ущерба в производительности самого канала связи.

Таблица 1.1 – Сравнение протоколов

Название протокола	VXLAN	STT	NVGRE
Инкапсуляция	– Использование UDP-протокола – UDP порт 8472 – Добавляется VXLAN заголовок размером 8 байт – Инкапсуляция IP и не-IP кадров	– Использование TCP-протокола – Добавляется STT заголовок размером 8 байт – Использование нестандартного TCP – Инкапсуляция IP и не-IP кадров	– Использование GRE-инкапсуляции – Использование GRE номера протокола 0x6558 – Инкапсуляция нетегированных IP и не-IP кадров
Идентификатор оверлейной сети	24-битный VNI	64-битный Context ID	24-битный VSID, плюс 8 бит идентификатор потока
Размер служебных заголовков	50 байт	76 байт	42 байта
ECMP и балансировка в порт-группе	– Порт источника в VXLAN заголовке в результате хеширования внутренних заголовков – Транспорт должен поддерживать хэширование на базе 5 параметров (src+dst ip, protocol, src+dst port)	– Порт источника в STT заголовке в результате хеширования заголовков – Транспорт должен поддерживать хэширование на базе 5 параметров (src+dst ip, protocol, src+dst port)	– Черновик RFC предлагает 32-битный VSID + 8-битный flow ID – Хеширование GRE не предполагается на аппаратных устройствах
Распространение информации о подключенных узлах	– Flood+Learn – MP-BGP EVPN	На базе OpenFlow	Использование любого механизма (чтение данных из транспорта, использование центрального сервера)
Поддержка виртуальными коммутаторами	Cisco Nexus 1000v и VMware DVS	Nicira, Open vSwitch	Microsoft Hyper-V virtual switch
Масштабируемость	1 млн хостов	1000 хостов	Неизвестно
Поддержка производителями	Cisco, VMware, Arista, Brocade, Citrix, Red Hat, Broadcom	VMware, Broadcom	Microsoft, Arista, Emulex, Huawei, HP
Шлюзы между виртуальными и аппаратными сетями	VMware vShield, Cisco ASA for Nexus 1000V Series Switch, Cisco Cloud Services Router (CSR) 1000V Series, Arista 7150 switch, Brocade ADX	Решение Nicira	Неизвестно
Сервисные цепочки	Cisco vPath в Cisco Nexus 1000v	Неизвестно	Неизвестно
Стандарт	RFC 7348 + RFC 7432	draft-davie-stt-08	RFC 7637

По состоянию на июнь 2017 протокол GENEVE находится в стадии утверждения спецификаций, поэтому на рынке не представлены готовые решения на его основе [3].

Результаты сравнения текущих версий оверлейных протоколов представлены в таблице 1.1.

2 Создание оверлейной схемы в виртуальной среде EVE-NG

Для реализации топологии был использован бесплатный продукт EVE-NG (старое название Unetlab) [4].

Физическая топология представляет собой звезду, центром которой является vHUB, к которому подключены все сетевые устройства – как уровня провайдера (PE, P), так и уровня клиента (CE). Используемая физическая топология представлена на рисунок 2.1.

vHUB, как и следует из названия, транслирует все приходящие пакеты во все порты, кроме того, от которого он их получил. Это позволяет прослушивать все проходящие пакеты в топологии будучи подключенным к порту любого из устройств. Для этого используется программный пакет Wireshark. Также это позволяет избавиться от дополнительной настройки коммутатора, чтобы он мог успешно передавать все тегируемые пакеты.

В данной схеме будет реализовано подключение CE-устройств (в виде CSR1000v) к PE (IOL), а также показаны настройки и отсутствие особых требований к транзитной сети, т.к. иначе возможны дополнительные финансовые траты за предоставление услуг сторонним провайдером. Из требований к провайдеру есть лишь предоставление unicast-доступности, а также поддержка BiDir-PIM.

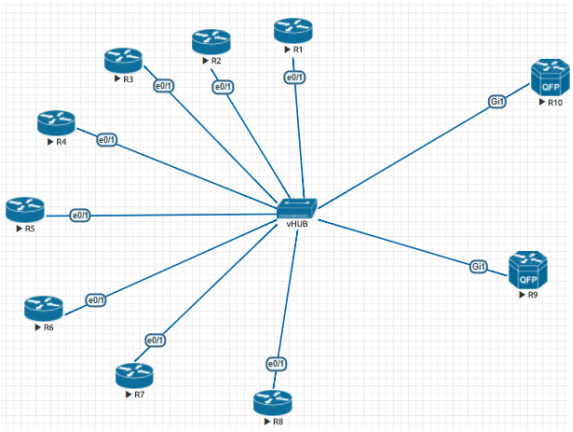

Рисунок 2.1 – Физическая топология

Логическая топология ISP BBB представлена на рисунке 2.2.

Возможно отказаться от использования двунаправленного мультикаст-протокола, в этом случае необходимо указывать расположение всех VTEP-шлюзов, а репликация всего трафика будет происходить на входе, когда один пакет копируется и посылается каждому шлюзу персонально. Это увеличивает нагрузку на сетевую инфраструктуру, но позволяет уйти от требований к наличию мультикаст-протоколов [5].

Клиентские устройства CE (Customer Edge) R9 и R10 будут подключены к PE (Provider Edge) R7 и R8. В качестве RR (Route-Reflector) для протокола BGP будет использован R1. Он будет выполнять Best-Path Selection алгоритм и отдавать своим RR-клиентам NLRI+next-hop+метрику. Хоть RR и скрывает истинную картину путей, он позволяет не создавать избыточных

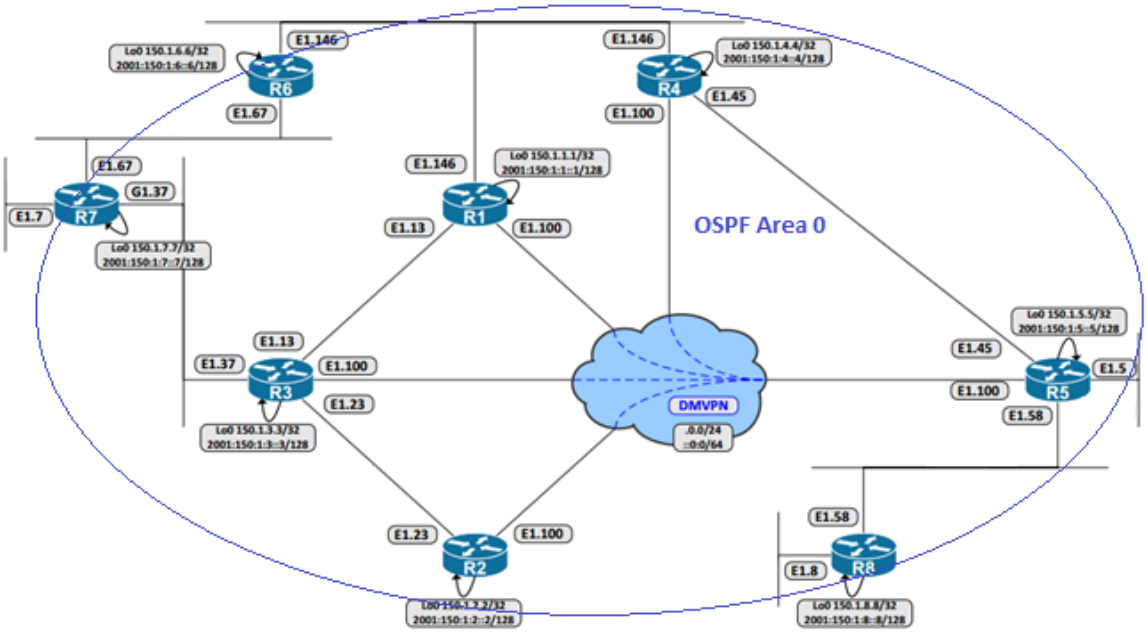

Рисунок 2.2 – Логическая топология оператора связи

Prevention Mechanism для iBGP-соседств. Суть iBGP Loop Prevention Mechanism заключается в том, что маршрут, принятый от iBGP-соседа не анонсируется другому iBGP-соседу. Также в случае с RR не действует другое правило BGP – маршрутизатор не принимает апдейты, если он не использует Route-Target, описываемый в атрибутах NLRI.

Сама же внутренняя сеть ISP BBB будет состоять из единого OSPF-домена. Также, исходя из небольшого размера сети и количества устройств, все они будут помещены в одну зону, для упрощения. На рисунке 2.3 видно OSPF + MPLS топологию провайдера BBB. В данном случае MPLS будет автоматически распространён средствами OSPF, а именно включением функции «mpls ldp autoconfig». Это макрос, включающий анонсирование служебной информации на интерфейсах, где работает OSPF-процесс. В некоторых случаях можно вручную включать MPLS-процесс на интерфейсе, в этом случае предлагается дополнительная защита от человеческого фактора, когда ошибка в настройке будет весьма ощутима для инфраструктуры [6].

Для банка сегмент, предоставляющий VXLAN-сервис, будет выглядеть следующим образом (рисунок 2.4).

Топологию, которая будет настраиваться, можно увидеть на рисунке 2.5. В качестве PC1 и PC2 будет временно использоваться L2-IOL как сетевая аналогия PC, так как интересует техническая часть.

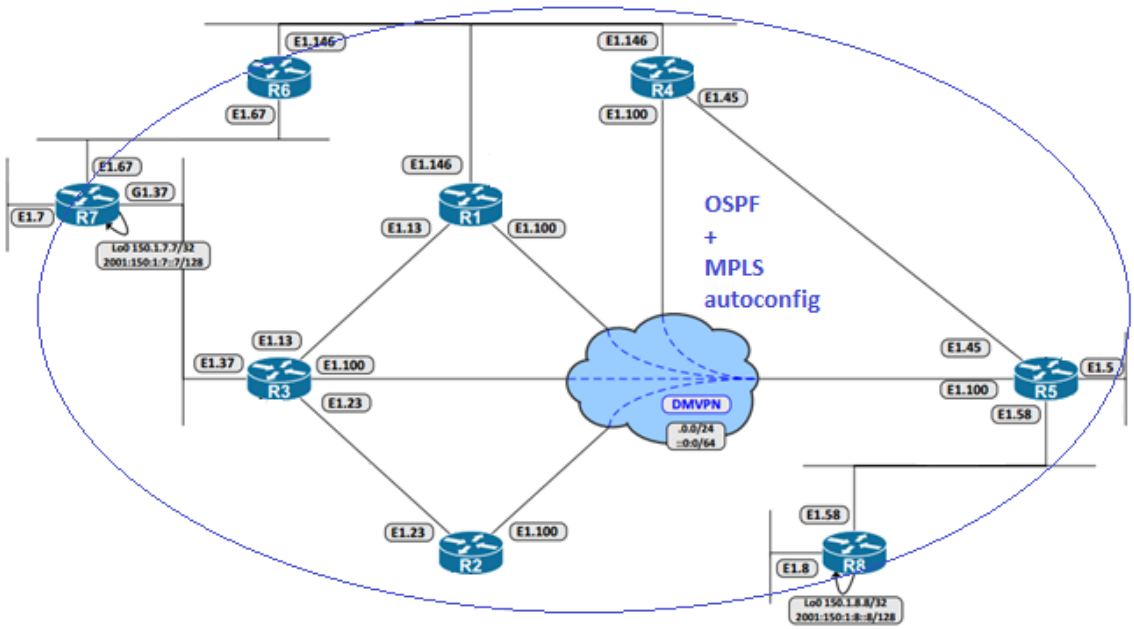

Рисунок 2.3 – Топология MPLS-облака оператора связи

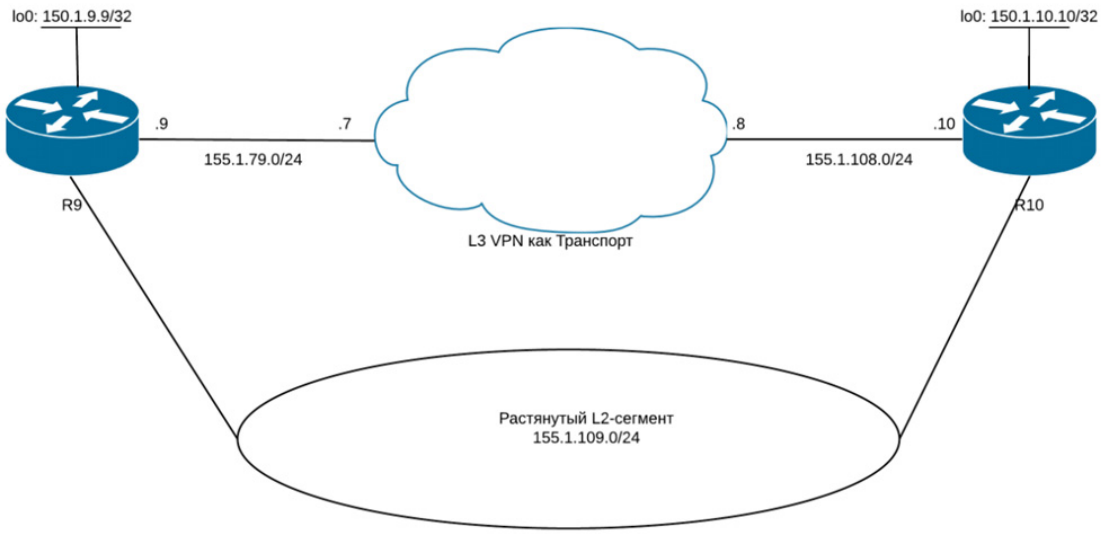

Рисунок 2.4 – Упрощённая схема подключения растянутого L2-сегмента

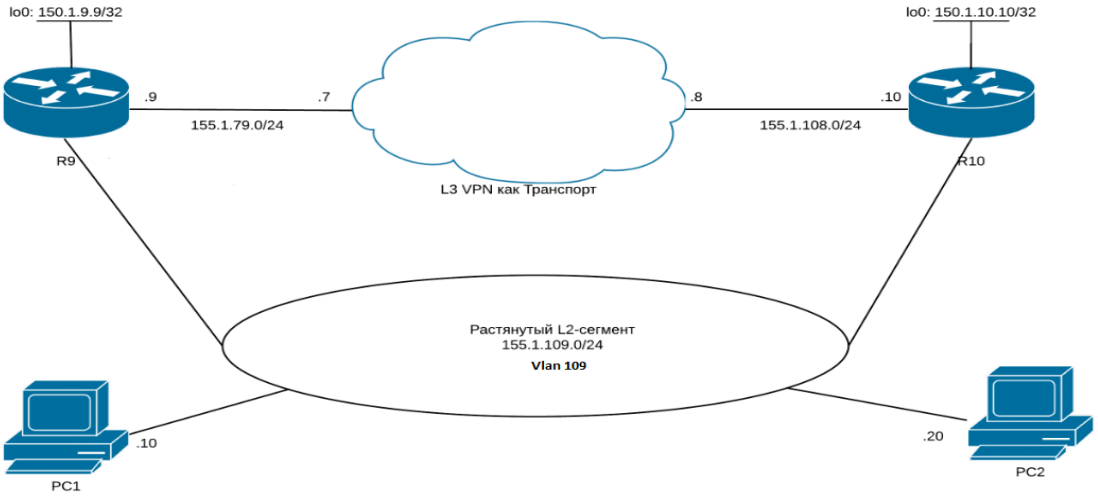

Рисунок 2.5 – Схема растягиваемой сети

3 Конфигурация VXLAN-оверлея

Конфигурация VXLAN-шлюза представлена на рисунке 3.1. В качестве моста между L2-сегментом и L3-транспортом выступает интерфейс NVE1, при его конфигурации указывается интерфейс-источника, привязка VNI (идентификатор расширенного влана) к мультикаст-группе, о которой данное устройство будет сообщать в PIM-Join пакетах в сторону RP (Rendezvous Point – точка к которой строят SharedTree все PIM-устройства). В этой конфигурации указывается, какой именно VLAN «растягивается», что позволяет растянуть большое количество виртуальных сетей (максимальное значение очень зависит от уровня оборудования). Теперь проверим состояние NVE-туннеля (рисунки 3.2, 3.3).

По-умолчанию используется порт UDP 4789 (рисунок 3.4), но его можно при необходимости изменить на любой другой, но делать это необходимо сразу на всех VTEP-устройствах.

Размер передаваемого трафика зависит от MTU на всём пути следования, поэтому для поддержки Jumbo-frame необходимо настроить MTU на соответствующее значение на всех устройствах сети.

Список переданных пакетов можно увидеть в packet-capture раздела Wireshark (рисунок 3.5).

Wireshark автоматически «разворачивает» внешние заголовки и показывает внутренние данные, поэтому можно увидеть детали пакета (рисунок 3.6). На нём представлена полная структура VXLAN-пакета – полный Layer 2 кадр завёрнут в VXLAN-заголовок, а дальше инкапсулирован в UDP-сегмент. В адресной информации можно заметить порт получателя (4789, который возможно изменить на любой другой), в качестве адресов отправителя и получателя можно увидеть адреса лупбэков VTEP-устройств. Вывод данного пакета косвенно подтверждает всю конфигурацию, которая была применена на настраиваемых виртуальных устройствах.

На этом локальная настройка узлов сети завершена.

```
R9#sh run | sec Loopback|nve|bridge-domain|GigabitEthernet2|bidir
bridge-domain 109
 member vni 5000
 member GigabitEthernet2 service-instance 109
interface Loopback0
 ip address 150.1.9.9 255.255.255.255
 ip pim sparse-mode
interface nve1
 no ip address
 member vni 5000 mcast-group 225.1.1.1
 source-interface Loopback0
interface GigabitEthernet2
 description "to STRETCHED L2"
 no ip address
 negotiation auto
 service instance 109 ethernet
  encapsulation dot1q 109
  rewrite ingress tag pop 1 symmetric
 !
ip pim bidir-enable
ip pim rp-address 100.100.100.100 bidir
R9#
```

Рисунок 3.1 – VXLAN конфигурация на VTEP R9

```
R9#
R9#show nve interface nve 1 detail
Interface: nve1, State: Admin Up, Oper Up Encapsulation: Vxlan
source-interface: Loopback0 (primary:150.1.9.9 vrf:0)
   Pkts In   Bytes In   Pkts Out  Bytes Out
        95     118008        117     145952
R9#
```

Рисунок 3.2 – Информация о транслированных пакетах

```
R9#show nve peers
Interface  Peer-IP          VNI        Peer state
   nve1    150.1.10.10      5000       -
R9#
```

Рисунок 3.3 – Вывод VTEP-соседств

```
R9#
R9#show platform software vxlan F0 udp-port
VXLAN UDP Port: 4789

R9#
```

Рисунок 3.4 – Вывод используемого порта для обмена служебной информации

```
14 12:56:38.803595  155.1.109.10  155.1.109.20  ICMP  1468 Echo (ping) request  id=0x000b, seq=0/0, ttl=255 (reply in 15)
15 12:56:38.809014  155.1.109.20  155.1.109.10  ICMP  1468 Echo (ping) reply    id=0x000b, seq=0/0, ttl=255 (request in 14)
16 12:56:38.868598  155.1.109.10  155.1.109.20  ICMP  1468 Echo (ping) request  id=0x000b, seq=1/256, ttl=255 (reply in 17)
17 12:56:38.987031  155.1.109.20  155.1.109.10  ICMP  1468 Echo (ping) reply    id=0x000b, seq=1/256, ttl=255 (request in 16)
18 12:56:39.104270  155.1.109.10  155.1.109.20  ICMP  1468 Echo (ping) request  id=0x000b, seq=2/512, ttl=255 (reply in 19)
19 12:56:39.217694  155.1.109.20  155.1.109.10  ICMP  1468 Echo (ping) reply    id=0x000b, seq=2/512, ttl=255 (request in 18)
20 12:56:39.340690  155.1.109.10  155.1.109.20  ICMP  1468 Echo (ping) request  id=0x000b, seq=3/768, ttl=255 (reply in 21)
21 12:56:39.448845  155.1.109.20  155.1.109.10  ICMP  1468 Echo (ping) reply    id=0x000b, seq=3/768, ttl=255 (request in 20)
22 12:56:39.569274  155.1.109.10  155.1.109.20  ICMP  1468 Echo (ping) request  id=0x000b, seq=4/1024, ttl=255 (reply in 23)
23 12:56:39.688860  155.1.109.20  155.1.109.10  ICMP  1468 Echo (ping) reply    id=0x000b, seq=4/1024, ttl=255 (request in 22)
```

Рисунок 3.5 – Вывод переданных пакетов посредством VXLAN

```
▷ Frame 22: 1468 bytes on wire (11744 bits), 1468 bytes captured (11744 bits) on interface 0
▷ Ethernet II, Src: 50:00:00:01:00:00 (50:00:00:01:00:00), Dst: aa:bb:cc:00:02:00 (aa:bb:cc:00:02:00)
▷ 802.1Q Virtual LAN, PRI: 0, CFI: 0, ID: 79
▷ Internet Protocol Version 4, Src: 150.1.9.9, Dst: 150.1.10.10
◢ User Datagram Protocol, Src Port: 28615, Dst Port: 4789
    Source Port: 28615
    Destination Port: 4789
    Length: 1430
    [Checksum: [missing]]
    [Checksum Status: Not present]
    [Stream index: 2]
◢ Virtual eXtensible Local Area Network
  ▷ Flags: 0x0800, VXLAN Network ID (VNI)
    Group Policy ID: 0
    VXLAN Network Identifier (VNI): 5000
    Reserved: 0
▷ Ethernet II, Src: aa:bb:cc:80:04:00 (aa:bb:cc:80:04:00), Dst: aa:bb:cc:80:05:00 (aa:bb:cc:80:05:00)
▷ Internet Protocol Version 4, Src: 155.1.109.10, Dst: 155.1.109.20
▷ Internet Control Message Protocol
```

Рисунок 3.6 – Структура инкапсулированного пакета

Заключение

В статье рассмотрены принципы и технологии построения логических каналов L2 over L3 уровня для организации бесшовной связи удаленных филиалов организации. В результате применения технологии создана модель программно-аппаратной схемы, которая позволяет передавать данные на любое расстояние без дополнительных требований к промежуточному оборудованию.

ЛИТЕРАТУРА

1. *Демиденко*, *О.М.* Принципы формализации вычислительного процесса в ЛВС / О.М. Демиденко // Известия Гомельского государственного университета им. Ф. Скорины. – 2017. – № 6 (105). – С. 75–78.

2. *Демиденко*, *О.М.* Уровни представления вычислительного процесса и рабочей нагрузки на ЛВС / О.М. Демиденко // Цифровая трансформация. – 2017. – № 1. – С. 11–15.

3. *Репозиторий информационных документов* [Электронный ресурс]. – Режим доступа: https://www.ietf.org/. – Дата доступа: 24.06.2017.

4. *Демиденко*, *О.М.* Изучение влияния внешних помех на качество сигнала в сетях WI-FI / О.М. Демиденко, В.Н. Кулинченко // Проблемы физики, математики и техники. – 2015. – № 4 (25). – С. 96–99.

5. *Демиденко*, *О.М.* Функциональные возможности программного комплекса адаптивной идентификации пользователей корпоративной сети / О.М. Демиденко, В.Д. Левчук, А.И. Кучеров // Проблемы физики, математики и техники. – 2010. – № 3 (4). – С. 69–73.

6. *Архитектура программного инструментария по обеспечению надежности узла ЛВС* / А.И. Кучеров, А.В. Воруев, О.М. Демиденко, В.Д. Левчук // Проблемы физики, математики и техники. – 2017. – № 4 (33). – С. 100–103.

Поступила в редакцию 07.02.18.