Internet Engineering Task Force (IETF)
Request for Comments: 8239
Category: Informational
ISSN: 2070-1721

L. Avramov
Google
J. Rapp
VMware

August 2017

Методология оценки производительности ЦОД

Data Center Benchmarking Methodology

Тезисы

Целью этого информационного документа является описание методологии и методов измерений для оценки производительности сетевого оборудования ЦОД. В предшествующем документе RFC 8238 описана терминология, которая считается нормативной. Многие из этих терминов и методов могут применяться к сетевому оборудованию, не относящемуся к ЦОД, поскольку разработанные для таких центров технологии могут применяться в других местах.

Статус документа

Документ не является спецификацией стандарта (Internet Standards Track) и публикуется с информационными целями.

Документ является результатом работы IETF[1] и представляет согласованный взгляд сообщества IETF. Документ прошел открытое обсуждение и был одобрен для публикации IESG[2]. Не все одобренные IESG документы претендуют на статус Internet Standard (см. раздел 2 в RFC 7841).

Информацию о текущем статусе документа, ошибках и способах обратной связи можно найти по ссылке http://www.rfc-editor.org/info/rfc8239.

Оглавление

[1]Internet Engineering Task Force.
[2]Internet Engineering Steering Group.

1. Введение

Картина трафика в ЦОД неоднородна и постоянно меняется. Это обусловлено характером и разнообразием применяемых в ЦОД приложений. Это могут быть большие потоки «запад-восток» («горизонтальные» потоки между серверами одного ЦОД) в одном центре и большие потоки «север-юг» («вертикальные» потоки из внешнего источника к серверу ЦОД) в другом, а также разные комбинации направлений потоков. Картины трафика по своей природе содержат пики (всплески) и включают потоки «многие к одному» «многие ко многим», «один ко многим». Потоки могут быть небольшими и чувствительными к задержками или большими и чувствительными к пропускной способности, а также включать смесь трафика UDP и TCP. Все перечисленное может существовать в одном кластере и поток может проходить через одно сетевое устройство. Тесты производительности сетевых устройств используются достаточно давно и описаны в [RFC1242], [RFC2432], [RFC2544], [RFC2889] и [RFC3918]. Эти тесты в основном привязаны к параметрам задержки и максимальной пропускной способности [RFC2889] тестируемого устройства (DUT[1]). Эти стандарты хороши для измерения теоретической максимальной пропускной способности, скорости пересылки и задержки в условиях теста, но не соответствуют реальным картинам трафика, который может проходить через сетевые устройства.

Ниже перечислены основные характеристики типичных сетевых устройств.

- Высокая плотность портов(не менее 48).
- Высокая скорость (вплоть до 100 Гбит/с на порт).
- Высокая пропускная способность (суммарная линейная скорость всех портов для уровня 2 и/или 3).
- Малые задержки (микросекунды или наносекунды).
- Незначительный объем буферов (мегабайты в объеме всего устройства).
- Пересылка на уровнях 2 и 3 (уровень 3 не обязателен).

В этом документе рассматривается методология оценки производительности физического сетевого оборудования (DUT) в составе ЦОД, включая сценарии с перегрузкой, анализ буферизации в коммутаторах, микропики, блокировку линий, а также разных ситуаций для смешанного трафика. В [RFC8238] приведены определения терминов, которые считаются нормативными для тестирования.

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), **не рекомендуется** (NOT RECOMMENDED), **возможно** (MAY), **необязательно** (OPTIONAL) в данном документе должны интерпретироваться в соответствии с BCP 14 [RFC2119] [RFC8174] тогда и только тогда, когда они набраны заглавными буквами (выделены **шрифтом**), как показано здесь.

1.2. Формат методологии и рекомендации по воспроизводимости

В разделах 2 - 6 этого документы используется деление на перечисленные ниже параграфы.

- Цели
- Методология
- Формат отчета

Для каждой методологии тестирования, описанной в этом документе, важнейшее значение имеет повторяемость тестов. Рекомендуется выполнять тесты неоднократно для обеспечения уверенности в результатах. Особенно важно это для тестов, описанных в разделе 3, поскольку тестирование буферизации исторически было наименее надежным. **Следует** указывать в отчете число повторов. **Следует** добиваться относительного стандартного отклонения ниже 10%.

2. Тестирование скорости в линии

2.1. Цели

Целью этого метода является тестирование «максимальной скорости» (maximum rate) для пропускной способности, задержки и ее вариаций (jitter). Целями являются (1) выполнение теста (2) методология проверки способности DUT пересылать пакеты со скоростью среды при отсутствии насыщения.

2.2. Методология

Генератор трафика **следует** подключать ко все портам DUT. **Должны** выполняться два типа тестов: (1) тестирование пар портов [RFC2544] [RFC3918] и (2) тестирование с использованием полносвязной топологии (full-mesh) [RFC2889] [RFC3918].

Для всех тестов скорость передачи генератора трафика **должна** быть не более 99,98% от номинальной скорости линии (без дополнительной настройки PPM с учетом отклонений частоты на интерфейсах) для создания для устройства DUT «разумно жесткой» нагрузки (см. раздел 5 в [RFC8238]). **Можно** также представить результаты теста на меньшей скорости для лучшего понимания роста производительности в части задержек и их вариаций при скорости линии ниже 99,98%. Скорость поступления трафика **следует** измерять в процессе тестирования в процентах от скорости линии.

Тест **должен** обеспечивать статистику результатов - как минимум, максимальное, среднее и минимальное значение задержки для одинаковых итераций теста.

Тест **должен** обеспечивать статистику результатов - как минимум, максимальное, среднее и минимальное значение вариаций задержки для одинаковых итераций теста.

[1]Device Under Test.

В дополнение к этому для случаев, когда генератор трафика не может быть подключен ко всем портам DUT, **должен** использоваться snake-тест для тестирования скорости линий с исключением задержки и ее вариаций, как не имеющих отношения к делу. Описание этого теста приведено ниже.

- Генератор трафика подключается к первому и последнему порту DUT.
- Порты попарно соединяются между собой в цепочку («змейку» - snake) - порт 2 с портом 3, порт 4 с портом 5 и т. д., порт N-2 с портом N-1, где N - общее число портов DUT.
- Порты 1 и 2 помещаются в одну VLAN X, порты 3 и 4 - в VLAN Y и т. д, порты N-1 и N - в одну VLAN Z.

Этот тест позволяет проверить скорость линии для уровней 2 и 3 [RFC2544] [RFC3918] в тех случаях, когда генератор трафика имеет лишь два порта. Задержки и их вариации в этом тесте не проверяются.

2.3. Формат отчета

Отчет должен включать перечисленные ниже сведения.

- Данные калибровки на физическом уровне в соответствии с разделом 4 [RFC8238].
- Число использованных портов.
- Скорость приема в процентах от пропускной способности при скорости передачи 99,98% от номинальной скорости линии на каждом порту для всех размеров пакетов от 64 до 9216 байтов. Рекомендуется увеличивать размер пакетов на 64 байта для каждой итерации, но приемлемы также размеры шага в 256 и 512 байтов. Чаще всего в отчетах указывают данные для пакетов размером 64, 128, 256, 512, 1024, 1518, 4096, 8000 и 9216 байтов.

Схему тестирования можно сформулировать с использованием [RFC6985].

- Пропускная способность должна представляться в процентах от общего числа переданных кадров.
- Отбрасывание пакетов **должно** представляться число отброшенных пакетов, **следует** также представлять его в процентах от скорости линии.
- Значения задержки и ее вариаций указываются в единицах времени (обычно в миллисекундах или наносекундах) для пакетов размером от 64 до 9216 байтов.
- Для задержки и ее вариаций указываются минимальное, среднее и максимальное значения. Если для получения минимального, среднего и максимального значений используются разные итерации, это **следует** отметить в отчете вместе с указанием причин, не позволивших получить все три значения в одной итерации теста.
- Для вариаций задержки **рекомендуется** представлять также гистограмму, показывающую распределение числа с разными значениями задержки.
- Тесты пропускное способности, задержки и ее вариаций **можно** проводить независимо с предоставлением в отчете надлежащей документации, однако **следует** проводить эти тексты одновременно.
- Методология предполагает наличие в устройстве DUT не менее 9 портов, поскольку для некоторых тестов нужны 9 и более портов.

3. Тестирование буферов

3.1. Цели

Задачей этого теста является измерение размера DUT буферов в типичных/разных/многочисленных ситуациях. Архитектура буферизации в разных DUT может различаться и включать буферизацию на выходе, общий выходной буфер SoC[1] (однокристальный коммутатор), буферизацию на входе или комбинацию перечисленных вариантов. Методология тестирования обеспечивает измерение буферов независимо от используемой в DUT архитектуры буферизации.

3.2. Методология

Генератор трафика **должен** подключаться ко всем портам DUT. Методология оценки буферизации для коммутаторов ЦОД основана на использовании информации о перегрузке пакетами известного размера и результатах измерений времени задержки. Максимальная задержка будет возрастать, пока не будет отброшен первый пакет. С этого момента максимальная задержка будет сохраняться. Это определяет точку перегиба кривой максимальной задержки (переход к горизонтальному участку). Множество входных портов **должно** получать известное число кадров фиксированного (известного) размера, адресованных в один выходной порт, для создания понятной перегрузки. Общее число пакетов, переданных из порта, обеспечивающего избыточный трафик, за вычетом одного, представляет максимальный размер буфера порта в точке перегиба.

В описании представленных ниже процедур 1), 2), 3) и 4) используются «первая итерация», «вторая итерация» и «последняя итерация». Идея этого состоит в том, чтобы показать читателю логику изменений параметров теста в каждой итерации. Последняя итерация показывает финальное состояние переменных.

1) Максимальная эффективность буферизации.
   - Первая итерация. Входной порт 1 передает 64-байтовые пакеты со скоростью линии в выходной порт 2, а порт 3 передает небольшое известное количество избыточного (oversubscription) трафика (рекомендуется 1%) таких же пакетов (64 байта) в выходной порт 2. Размер буфера определяется произведением размера кадра на число кадров избыточного трафика в точке перегиба.

[1]Switch-on-Chip - микросхема коммутации.

- Вторая итерация. Входной порт 1 передает 65-байтовые пакеты со скоростью линии в выходной порт 2, а порт 3 передает небольшое известное количество избыточного трафика (рекомендуется 1%) таких же пакетов (65 байтов) в выходной порт 2. Размер буфера определяется произведением размера кадра на число кадров избыточного трафика в точке перегиба.

- Последняя итерация. Входной порт 1 передает пакеты размеров B байтов со скоростью линии в выходной порт 2, а порт 3 передает небольшое известное количество избыточного трафика (рекомендуется 1%) таких же пакетов (B байтов) в выходной порт 2. Размер буфера определяется произведением размера кадра на число кадров избыточного трафика в точке перегиба.

- Когда станет ясно, что значение B определяет максимальный размер буфера, значение B служит для определения максимальной эффективности буферизации.

2) Измерение максимального размера буфера порта.

   При фиксированном размере пакетов B как определено в процедуре 1) для фиксированного, принятого по умолчанию значения DSCP[1]/CoS[2] = 0 и трафике с индивидуальной адресацией выполняются приведенные ниже процедуры.

   - Первая итерация. Входной порт 1 передает со скоростью линии в выходной порт 2, тогда как порт 3 передает известное небольшое количество избыточного трафика (рекомендуется 1%) с таким же размером пакетов в выходной порт 2. Измеряется размер буфера путем умножения числа переданных избыточных кадров на размер кадра.

   - Вторая итерация. Входной порт 2 передает со скоростью линии в выходной порт 3, тогда как порт 4 передает известное небольшое количество избыточного трафика (рекомендуется 1%) с таким же размером пакетов в выходной порт 3. Измеряется размер буфера путем умножения числа переданных избыточных кадров на размер кадра.

   - Последняя итерация. Входной порт N-2 передает со скоростью линии в выходной порт N-1, тогда как порт N передает известное небольшое количество избыточного трафика (рекомендуется 1%) с таким же размером пакетов в выходной порт N-1. Измеряется размер буфера путем умножения числа переданных избыточных кадров на размер кадра.

   Эту последовательность тестов **можно** повторять с использованием разных значений DSCP/CoS для трафика, а затем использовать групповой трафик для проверки влияния DSCP/CoS на размер буфера.

3) Измерение максимальных размеров буферов пары портов.

   - Первая итерация. Входной порт 1 передает со скоростью линии в выходной порт 2, входной порт 3 - в выходной порт 4 и т. д. Входные порты N-1 и N будут давать избыточный трафик в размере 1% от скорости линии в выходные порты 2 и 3, соответственно. Значение размера буфера измеряется путем умножения числа переданных избыточных кадров на размер кадра для каждого выходного порта.

   - Вторая итерация. Входной порт 1 передает со скоростью линии в выходной порт 2, входной порт 3 - в выходной порт 4 и т. д. Входные порты N-1 и N будут давать избыточный трафик в размере 1% от скорости линии в выходные порты 4 и 5, соответственно. Размер буфера измеряется путем умножения числа переданных избыточных кадров на размер кадра для каждого выходного порта.

   - Последняя итерация. Входной порт 1 передает со скоростью линии в выходной порт 2, входной порт 3 - в выходной порт 4 и т. д. Входные порты N-1 и N будут давать избыточный трафик в размере 1% от скорости линии в выходные порты N-3 и N-2, соответственно. Значение размера буфера измеряется путем умножения числа переданных избыточных кадров на размер кадра для каждого выходного порта.

   Эта серия тестов **может** повторяться с использованием разных значений DSCP/CoS, а также с использованием группового трафика.

4) Измеряется максимальный размер буфера DUT с портами «множество в один».

   - Первая итерация. Входные порты 1,2,... N-1 передают со скоростью [(1/[N-1])*99.98]+[1/[N-1]]% от скорости линии в выходной порт N.

   - Вторая итерация. Входные порты 2,... N передают со скоростью [(1/[N-1])*99.98]+[1/[N-1]]% от скорости линии в выходной порт 1.

   - Последняя итерация. Входные порты N,1,2...N-2 передают со скоростью [(1/[N-1])*99.98]+[1/[N-1]]% от скорости линии в выходной порт N-1.

Эта последовательность тестов **может** повторяться для разных CoS в трафике, а затем для группового трафика.

Следует использовать индивидуальный, а затем групповой трафик для определения доли буфера при документированном выборе тестов. Также **следует** изменять значения CoS в пакетах для каждой итерации тестов, поскольку размер выделяемых буферов **может** зависеть от CoS. **Рекомендуется** менять входные и выходные порты случайным образом (документируя это) в разных тестах, чтобы измерить размер буфера на каждом порту DUT.

3.3. Формат отчета

Отчет **должен** включать указанную ниже информацию.

- Размер пакетов, при котором обеспечивалось наиболее эффективное использование буферов, вместе с DSCP/CoS.

[1]Differentiated Services Code Point - код дифференцированного обслуживания.
[2]Class of Service - класс обслуживания.

- Максимальный размер буфера для каждого порта.
- Максимальный размер буфера DUT.
- Размер использованных для теста пакетов.
- Размер «переподписки», если он отличался от 1%.
- Число входных и выходных портов и из расположение в DUT.
- Нужно указать повторяемость тестов - число итераций одного теста и процент отклонения между результатами в этих итерациях (максимум, минимум, среднее).

Процент отклонения - это показатель, определяющий разницу между измеренным и предыдущими значениями.

Например, при тестировании задержки измеряется минимальное значение и процент отклонения (PV[1]) минимального значения будет показывать наскольку эта величина различается в текущем и предыдущем измерении.

PV = ((x2-x1)/x1)*100, где x2 - минимальное значение задержки в текущем тесте, а x1 - в предыдущем.

Такая же формула применяется для отклонений максимального и среднего значения.

4. Тестирование микропиков трафика

4.1. Цели

Целью этого теста является определение максимального числа микропиков пакетов, которое устройство DUT может выдержать при разных конфигурациях.

Этот тест обеспечивает методологию, дополняющую тесты, которые описаны в [RFC1242], [RFC2432], [RFC2544], [RFC2889] и [RFC3918].

- Все пики следует передавать с интенсивностью 100% (определена в параграфе 6.1.1 [RFC8238]).
- Для этого теста должны использоваться все порты DUT.
- Рекомендуется тестировать все порты одновременно.

4.2. Методология

Генератор трафика **должен** быть подключен ко всем портам DUT. Для того, чтобы вызвать перегрузку в два (или более) входных порта **должны** передаваться пики пакетов, адресованные в один выходной порт. Простейшим вариантом является передача из двух входных портов в один выходной (2 в 1).

Пики **должны** передаваться с интенсивностью (определена в параграфе 6.1.1 [RFC8238]) 100%, означающей передачу пакетов в рамках пика с минимальными межпакетными интервалами. Число пакетов в пике подбирается методом проб и увеличивается, пока не начнутся потери пакетов. Агрегатное число пакетов от всех отправителей используется для расчета максимального числа микропиков, которое способно выдержать устройство DUT.

Рекомендуется менять входные и выходные порты в разных тестах для измерения максимальной «емкости» микропиков.

Интенсивность микропиков (см. параграф 6.1.1 в [RFC8238]) **может** меняться для получения максимальной «емкости» при разных входных скоростях.

Рекомендуется тестировать все порты DUT одновременно и в разных конфигурациях для понимания всех комбинаций входных и выходных портов и интенсивностей.

Ниже приведен пример.

- Первая итерация. N-1 входных портов передают в один выходной порт.
- Вторая итерация. N-2 входных порта передают в два выходных порта.
- Последняя итерация. два входных порта передают в N-2 выходных порта.

4.3. Формат отчета

Отчет **должен** включать перечисленные ниже результаты.

- Максимальное число пакетов, полученных на входном порту с максимальным размером пика, при котором не наблюдается потерь.
- Размер пакетов, использованных при тестировании.
- Число входных и выходных портов, а также их расположение в DUT.
- Требуется описать повторяемость тестов - число итераций одного теста и величину отклонений между итерациями (максимум, минимум, среднее).

5. «Пробки»

5.1. Цели

Блокировка HOLB[2] или «пробка» - это влияющее на производительность явление, когда пакеты тормозятся первым пакетом, ожидающим пересылки в какой-то другой выходной порт. Определение этого феномена приведено в

[1]The percentage of variation.
[2]Head-of-line blocking.

параграфе 5.5 RFC 2889 (Congestion Control). Этот раздел служит расширением RFC 2889 в контексте оценки производительности ЦОД.

Цель этого теста заключается в определении поведения DUT в случаях HOLB и измерение уровня потери пакетов.

Различия между этим тестом HOLB и RFC 2889 перечислены ниже.

- Тест HOLB начинают с 8 портов в двух группах по 4 порта в каждой вместо тестирования на 4 портах, как в параграфе 5.5 RFC 2889.
- Во второй итерации теста HOLB номера всех портов сдвигают на 1, это тоже отличается от теста HOLB в RFC 2889. Сдвиг портов продолжается до тех пор, пока каждый порт не побывает в роли первого в группе - это делается для того, чтобы учесть все отклонения в поведении SoC в устройстве DUT.
- Другим отличием является увеличение числа портов в тесте HOLB, чтобы трафик распределялся между четырьмя портами вместо двух (25% на каждый порт вместо 50%).
- В параграфе 5.3 приведены требования, дополняющие требования параграфа 5.5 в RFC 2889.

5.2. Методология

Для инициирования «пробки» в форме HOLB применяется группа из 4 портов, два из которых являются входными, два - выходными. На первом входном порту **должны** быть настроены два потока, каждый из которых имеет свой выходной порт. Второй входной порт будет перегружать второй выходной порт, передавая данные со скоростью линии. Цель заключается в проверке наличия потерь потока для первого выходного порта, на котором нет переподписки.

Генератор трафика **должен** подключаться по меньшей мере к 8 портам DUT и **следует** подключать его ко всем портам.

Отметим, что тесты, описанные в процедурах 1) и 2) этого параграфа имеют итерации, называемые первой, второй и последней. Идея состоит в том, чтобы показать первые две итерации для понимания читателем логики процесса. Последняя итерация показывает конечное состояние переменных.

1) Измерения для двух групп с 8 портами DUT.
   - Первая итерация. Измеряется потеря пакетов для двух групп с последовательными портами.

     Операции для первой группы портов описаны ниже.

     Входной порт 1 передает 50% трафика в выходной порт 3 и 50% - в выходной порт 4. Входной порт 2 передает со скоростью линии в выходной порт 4. Измеряются потери для трафика из входного порта 1 в выходной порт 3.

     Операции для второй группы портов описаны ниже.

     Входной порт 5 передает 50% трафика в выходной порт 7 и 50% - в выходной порт 8. Входной порт 6 передает со скоростью линии в выходной порт 8. Измеряются потери для трафика из входного порта 5 в выходной порт 7.
   - Вторая итерация. Выполняется первая итерация со сдвигом всех портов на 1.

     Операции для первой группы портов описаны ниже.

     Входной порт 2 передает 50% трафика в выходной порт 4 и 50% - в выходной порт 5. Входной порт 3 передает со скоростью линии в выходной порт 5. Измеряются потери для трафика из входного порта 2 в выходной порт 4.

     Операции для второй группы портов описаны ниже.

     Входной порт 6 передает 50% трафика в выходной порт 8 и 50% - в выходной порт 9. Входной порт 7 передает со скоростью линии в выходной порт 9. Измеряются потери для трафика из входного порта 6 в выходной порт 8.
   - Последняя итерация. Первый порт первой группы подключен к последнему порту DUT, а последний порт второй группы подключен к седьмому порту DUT.

     Измеряются потери трафика из входного порта N в выходной порт 2 и из входного порта 4 ы выходной порт 6.
2) Измерения с N/4 групп для DUT с N портов.

   Трафик из входного порта расщепляется в 4 выходных порта (100/4 = 25%).
   - Первая итерация. Используются все порты DUT, выбираемые с шагом 4. Повторяется методология процедуры 1) для всех групп портов, доступных на устройстве и измеряются потери в каждой группе.
   - Вторая итерация. Номера портов в каждой группе смещаются на +1.
   - Последняя итерация. Номера портов в каждой группе смещаются на N-1 и измеряются потери для каждой группы.

5.3. Формат отчета

Для каждого теста отчет **должен** включать указанную ниже информацию.

- Конфигурация порта, включая число и расположение входных и выходных портов в DUT.
- Соответствие наблюдавшихся пробок описанию HOLB в разделе 5.
- Процент потерянного трафика.

- Требуется описать повторяемость тестов - число итераций одного теста и величину отклонений между итерациями (максимум, минимум, среднее).

6. Инкаст для трафика с учетом и без учета состояния

6.1. Цели

Целью этого теста является измерение значений TCP Goodput [TCP-INCAST] и задержки для комбинации больших и мелких потоков. Тест разработан для имитации смешанной среды потоков с учетом состояния (stateful flow), которым нужна высокая пропускная способность, и потоков без учета состояния, которые требуют малых задержек. Потоки с учетом состояния создаются путем генерации трафика TCP, а без учета состояния - путем генерации трафика UDP.

6.2. Методология

Для имитации воздействия трафика с учетом и без учета состояния на DUT **должно** использоваться множество входных портов, принимающих трафик, адресованный в один выходной порт. **Можно** также использовать комбинацию трафика с учетом и без учета состояния, приходящего в один входной порт. Простейшим вариантом будут два входных порта, принимающих трафик, который адресован в один выходной порт.

Один входной порт **должен** поддерживать через себя соединение TCP с получателем, подключенным к выходному порту. Трафик в потоке TCP **должен** передаваться с максимальной скоростью, поддерживаемой генератором трафика. Этот трафик TCP через DUT передается одновременно с трафиком без поддержки состояния, адресованным в тот же выходной порт. Трафик без поддержки состояния **должен** представлять собой микропик с интенсивностью 100%.

Рекомендуется менять входные и выходные порты в разных тестах, чтобы измерить максимальную «емкость» микропиков.

Интенсивность микропиков **можно** менять для получения «емкости» микропиков при разной входной скорости.

Рекомендуется использовать при тестировании все порты DUT.

Описанные здесь тесты имеют итерации, называемые первой, второй и последней. Идея состоит в том, чтобы показать первые две итерации для понимания читателем логики процесса. Последняя итерация показывает конечное состояние переменных.

Примеры

Вариации трафика с учетом состояний (TCP).

Для этого теста нужно генерировать трафик TCP. В итерациях теста число выходных портов **можно** менять.

- Первая итерация. Один входной порт получает трафик TCP с учетом состояния и один входной порт получает трафик без учета состояния, который передается в один выходной порт.
- Вторая итерация. Два входных порта получают трафик TCP с учетом состояния и один входной порт получает трафик без учета состояния, который передается в один выходной порт.
- Последняя итерация. N-2 входных портов получают трафик TCP с учетом состояния и один входной порт получает трафик без учета состояния, который передается в один выходной порт.

Вариации трафика без учета состояний (UDP).

Для этого теста нужно генерировать трафик UDP. В итерациях теста число выходных портов **можно** менять.

- Первая итерация. Один входной порт получает трафик TCP с учетом состояния и один входной порт получает трафик без учета состояния, который передается в один выходной порт.
- Вторая итерация. Один входной порт получает трафик TCP с учетом состояния и два входных порта получают трафик без учета состояния, который передается в один выходной порт.
- Последняя итерация. Один входной порт получает трафик TCP с учетом состояния и N-2 входных портов получают трафик без учета состояния, который передается в один выходной порт.

6.3. Формат отчета

Для каждого теста отчет **должен** включать указанную ниже информацию.

- Число входных и выходных портов с указанием привязки потоков с учетом и без учета состояния.
- Полезная пропускная способность потока с учетом состояния.
- Задержка трафика без учета состояния.
- Требуется описать повторяемость тестов - число итераций одного теста и величину отклонений между итерациями (максимум, минимум, среднее).

7. Вопросы безопасности

Измерения производительности, описанные здесь, ограничены характеристиками технологии в контролируемой лабораторной среде с выделенным адресным пространством и описанными выше ограничениями.

Топология тестовой сети должна быть независимой и **недопустимо** соединять ее с устройствами, которые могут пересылать тестовый трафик в работающие сети или в сеть управления тестированием.

Тесты производительности выполнялись по методу «черного ящика», когда все измерения проходят снаружи DUT.

В DUT **не следует** использовать при тестировании производительности нацеленные на это возможности. Любым влияния на безопасность сети со стороны DUT **следует** быть идентичными в тестовой и рабочей сети.

8. Взаимодействие с IANA

Этот документ не требует каких-либо действий со стороны IANA.

9. Литература

9.1. Нормативные документы

[RFC1242] Bradner, S., "Benchmarking Terminology for Network Interconnection Devices", RFC 1242, DOI 10.17487/RFC1242, July 1991, <https://www.rfc-editor.org/info/rfc1242>.

[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>.

[RFC2544] Bradner, S. and J. McQuaid, "Benchmarking Methodology for Network Interconnect Devices", RFC 2544, DOI 10.17487/RFC2544, March 1999, <https://www.rfc-editor.org/info/rfc2544>.

[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>.

[RFC8238] Avramov, L. and J. Rapp, "Data Center Benchmarking Terminology", RFC 8238, DOI 10.17487/RFC8238, August 2017, <https://www.rfc-editor.org/info/rfc8238>.

9.2. Дополнительная литература

[RFC2432] Dubray, K., "Terminology for IP Multicast Benchmarking", RFC 2432, DOI 10.17487/RFC2432, October 1998, <https://www.rfc-editor.org/info/rfc2432>.

[RFC2889] Mandeville, R. and J. Perser, "Benchmarking Metodology for LAN Switching Devices", RFC 2889, DOI 10.17487/RFC2889, August 2000, <https://www.rfc-editor.org/info/rfc2889>.

[RFC3918] Stopp, D. and B. Hickman, "Metodology for IP Multicast Benchmarking", RFC 3918, DOI 10.17487/RFC3918, October 2004, <https://www.rfc-editor.org/info/rfc3918>.

[RFC6985] Morton, A., "IMIX Genome: Specification of Variable Packet Sizes for Additional Testing", RFC 6985, DOI 10.17487/RFC6985, July 2013, <https://www.rfc-editor.org/info/rfc6985>.

[TCP-INCAST] Chen, Y., Griffith, R., Zats, D., Joseph, A., and R. Katz, "Understanding TCP Incast and Its Implications for Big Data Workloads", April 2012, <http://yanpeichen.com/professional/usenixLoginIncastReady.pdf>.

Благодарности

Авторы благодарят Al Morton и Scott Bradner за их рецензии и отклики.

Адреса авторов

Lucien Avramov

Google

1600 Amphitheatre Parkway

Mountain View, CA 94043

United States of America

Email: lucien.avramov@gmail.com

Jacob Rapp

VMware

3401 Hillview Ave.

Palo Alto, CA 94304

United States of America

Email: jhrapp@gmail.com

Перевод на русский язык

Николай Малых

nmalykh@gmail.com