

Network Working Group
Request for Comments: 3390
Obsoletes: 2414
Updates: 2581
Category: Standards Track

M. Allman
BBN/NASA GRC
S. Floyd
ICIR
C. Partridge
BBN Technologies
October 2002

Увеличение начального окна TCP

Increasing TCP's Initial Window

Статус документа

Документ содержит спецификацию стандартного протокола для сообщества Internet и является приглашением к дискуссии в целях развития и совершенствования протокола. Сведения о стандартизации и состоянии данного протокола можно найти в документе Internet Official Protocol Standards (STD 1). Допускается свободное распространение данного документа.

Авторские права

Copyright (C) The Internet Society (2002). All Rights Reserved.

Тезисы

Этот документ задаёт необязательный стандарт для расширения дозволенного начального окна TCP с 1 или 2 сегментов приблизительно до 4К байт, взамен требований RFC 2414. В документе рассматриваются преимущества и недостатки в результате увеличения начального размера окна, а также обсуждаются эксперименты и модели, показывающие, что увеличение начального размера окна не ведёт к коллапсу насыщения. В дополнение к этому документ содержит рекомендации для разработчиков.

Уровни требований

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

1. Изменение протокола TCP

Этот документ отменяет действие [RFC2414] и вносит изменения в [RFC2581], задавая увеличение верхнего предела начального размера окна TCP с 1 или 2 сегментов до 2 - 4 сегментов. В большинстве случаев это приводит к использованию начального размера окна приблизительно 4 кбайт (хотя с учётом разрешённого размера сегментов начальное окно размером в 2 сегмента может существенно превысить 4 кбайт).

Более точно верхнюю границу начального размера окна можно рассчитать с помощью выражения (1).

$$\min (4 * MSS, \max (2 * MSS, 4380 \text{ байт})) \quad (1)$$

Примечание. Передача пакета размером 1500 байтов говорит о максимальном размере сегмента (MSS) 1460 байтов (в предположении отсутствия опций IP и TCP). Следовательно, ограничение MSS для начального окна значением 4380 байтов позволяет отправителю в общем случае передать изначально три сегмента, используя для них пакеты размером 1500 байтов.

Если верхняя граница начального размера окна задаётся на основе MSS, возможны три варианта:

```
Если (MSS <= 1095 байтов)
    то win <= 4 * MSS;
если (1095 байтов < MSS < 2190 байтов)
    то win <= 4380;
Если (2190 байтов <= MSS)
    то win <= 2 * MSS;
```

Указанные значения начального размера окна являются рекомендуемыми - TCP **может** работать с окном большего размера. Однако мы предполагаем, что большинство реализаций TCP общего назначения выберет использование большего начального размера окна насыщения на основе уравнения (1).

Эти значения верхней границы начального окна отличаются от требований RFC 2581 [RFC2581], где сказано, что окно насыщения инициализируется с размером 1 или 2 сегмента.

Это изменение применяется к начальному окну соединения в течение первого периода кругового обхода (RTT¹) при передаче данных вслед за трехэтапным согласованием TCP. Ни в сегментах SYN/ACK, ни в подтверждениях (ACK) в процессе трехэтапного согласования размер начального окна не следует увеличивать сверх значения, заданного уравнением (1). Если сегмент SYN или SYN/ACK теряется, начальное окно, используемое отправителем после корректной передачи SYN **должно** быть равно 1 сегменту, содержащему MSS байтов.

¹Round trip time.

Реализации TCP используют процедуру замедленного старта в трёх разных ситуациях: (1) при старте нового соединения (начальное окно), (2) при восстановлении передачи после продолжительного бездействия (окно рестарта) и (3) при восстановлении передачи после тайм-аута повтора (окно потерь). Описанные в этом документе изменения оказывают влияние на начальное окно. Опционально TCP **может** установить для окна рестарта минимальное из значений начального окна и текущего `swnd` (иными словами, использование большего значения окна рестарта никогда не должно приводить к увеличению `swnd`). Предложенные здесь изменения **не** влияют на окно потерь, размер которого остаётся равным 1 сегменту, содержащему MSS байтов (чтобы обеспечить минимально возможный размер окна в случае некоторого насыщения).

2. Вопросы реализации

Когда увеличенное начальное окно реализовано вместе с Path MTU Discovery [RFC1191] и определено, что планируемое для использования значение MSS слишком велико, размер окна насыщения `swnd` **следует** уменьшить для предотвращения выброса большого числа мелких сегментов. Конкретно `swnd` **следует** уменьшить пропорционально снижению размера сегмента (поделить на отношение старого размера сегмента к новому размеру).

При реализации увеличенного начального окна вместе с Path MTU Discovery [RFC1191] возможна установка флага DF¹ во всех сегментах начального окна или в одном сегменте этого окна. Вопрос выбора лучшего из этих вариантов ещё не решён - будем надеяться, что опыт реализации поможет определиться с выбором. В первом случае, если начальные пакеты окажутся слишком велики, все эти пакеты будут отброшены в сети, поскольку их фрагментация запрещена флагом DF. Во втором случае, если пакеты окажутся слишком велики, все пакеты, кроме одного, будут фрагментированы в сети. Для второго случая установка флага DF в последнем сегменте начального окна обеспечивает минимальную вероятность повтора передачи по причине избыточного размера начального сегмента, поскольку в этом случае минимизируется вероятность доставки дубликатов ACK, включающей механизм Fast Retransmit. Однако вопросам взаимодействия увеличенного начального окна и механизма Path MTU Discovery требуется уделить дополнительное внимание.

Увеличенное начальное окно, описанное в этом документе, не предназначено для того, чтобы web-браузеры открывали множество одновременных соединений TCP в одном большом начальном окне. Когда браузер открывает одновременные соединения TCP с одним адресатом, он использует механизмы контроля насыщения TCP [FF99], независимо от размера начального окна. Использование этих механизмов вместе с увеличенным начальным окном оказывает дополнительное «давление» на прочий трафик в сети. Мы предлагаем использовать протокол HTTP/1.1 [RFC2068] (возобновляющиеся соединения TCP и конвейеры) в качестве средства повышения эффективности работы WEB.

3. Преимущества увеличения начального размера окна

1. Когда начальное окно имеет размер 1 сегмент, получатель, использующий отложенные подтверждения [RFC1122], вынужден ждать тайм-аута прежде, чем генерировать ACK. При начальном окне размером не менее 2 сегментов получатель будет генерировать ACK после прибытия второго сегмента данных. Это избавляет от необходимости ожидания тайм-аута (часто до 200 мсек и возможно до 500 мсек [RFC1122]).
2. Для соединений, передающих незначительный объем данных, увеличенное начальное окно снижает время передачи (в предположении невысокой вероятности отбрасывания пакетов). Для многих почтовых транзакций (SMTP [Pos82]) и просмотра web (HTTP [RFC1945, RFC2068]) объем передаваемой информации составляет менее 4К байт и увеличенный размер начального окна будет снижать время передачи данных до одного периода RTT.
3. Для соединений, которые способны использовать большое окно насыщения, предлагаемое изменение позволяет сэкономить до 3 периодов RTT и тайм-аута отложенных подтверждений в фазе замедленного старта. Это будет особенно полезно для широкополосных соединений со значительными задержками (например, через спутниковый канал).

4. Недостатки увеличения начального размера окна для соединения

В загруженных средах, особенно на маршрутизаторах с высоким уровнем насыщения, в частности, для маршрутизаторов с предубеждением к пикам трафика (как в типичных очередях Drop Tail), для соединения TCP может оказаться лучше использовать начальное окно размером в один сегмент. Существуют ситуации, когда соединение TCP, использующее замедленный старт с начальным окном размером в 1 сегмент, может обойтись без отбрасывания сегментов, а соединение с начальным окном в 4 сегмента может столкнуться с необходимостью повтора передачи по причине неспособности маршрутизатора обрабатывать незначительные пики трафика. Это может приводить к ненужным повторам передачи по тайм-ауту. Для соединений с большим окном, способным к восстановлению без тайм-аута повторной передачи, это может приводить к слишком раннему переходу от фазы замедленного старта к фазе увеличения окна. Такое преждевременное отбрасывание сегментов может возникать в незагруженных сетях с достаточной ёмкостью буферов или в сетях со средним уровнем насыщения, где перегруженные маршрутизаторы используют активное управления очередями (например, Random Early Detection [FJ93, RFC2309]).

Соединения TCP могут получать преимущества от увеличения начального размера окна даже если это увеличение приведёт к преждевременному отбрасыванию сегмента. Это будет наблюдаться, если (1) соединение TCP восстанавливается после отбрасывания сегмента без тайм-аута повторной передачи и (2) размер окна насыщения для соединения TCP ограничен малым значением по причине перегрузки в сети или малого окна, анонсируемого получателем.

5. Недостатки увеличения начального размера окна для сети

Для случаев потенциальной перегрузки будем отдельно рассматривать две потенциально опасных для сети ситуации. Первая связана со случаем, когда большое число сегментов перегруженного канала являются дубликатами сегментов, которые уже были получены приёмной стороной. Вторая ситуация возникает, когда большое число сегментов в перегруженных каналах будет позднее отброшено до поступления их конечному получателю.

¹Don't Fragment - не фрагментировать.

В плане негативного влияния на остальной трафик в сети потенциальным недостатком увеличенного начального окна будет то, что в этом случае увеличивается общая частота отбрасывания пакетов в сети. Ниже все эти негативные факторы рассматриваются более подробно.

Дубликаты сегментов

Как отмечено выше увеличенное начальное окно может иногда приводить к отбрасыванию сегмента из этого окна, хотя этот сегмент мог бы не оказаться отброшенным, если бы отправитель начал процедуру замедленного старта с окном размером в 1 сегмент. Однако в Приложении А показано, что даже в таких случаях увеличение начального размера окна не будет приводить к передаче большого числа дубликатов сегментов.

Сегменты, которые будут отброшены впоследствии

Насколько увеличение размера начального окна для TCP будет увеличивать в перегруженных каналах число сегментов, которые будут отброшены на пути к получателю? Эта проблема возникает только для соединений с множеством перегруженных каналов, где некоторые сегменты могут занимать дефицитную полосу в первом перегруженном канале на пути и потом оказаться отброшенными на следующих этапах доставки.

Во-первых, многие соединения TCP включают единственный перегруженный канал на пути через сеть. Отброшенные сегменты таких соединений не будут впустую «поглощать» дефицитную полосу пропускания и не будут вносить вклада в возникновение коллапса насыщения.

Однако некоторые пути могут включать множество перегруженных каналов и отбрасываемые сегменты начального окна могут потреблять полосу первых перегруженных каналов, а потом отбрасываться на последующих перегруженных участках пути. Поскольку рост скорости отбрасывания пакетов не зависит от размера начального окна TCP, проблема отбрасывания сегментов, которые уже прошли через перегруженные каналы до их отбрасывания в сети, будет одинакова для соединений TCP, независимо от того, будет начальное окно включать 4 сегмента или 1.

Рост частоты отбрасывания пакетов

Для сети с высокой частотой отбрасывания сегментов увеличение размера начального окна TCP может приводить к дополнительному росту частоты отбрасывания сегментов. Это связано отчасти с тем, что маршрутизаторы с механизмом управления очередями Drop Tail сталкиваются с трудностями при возникновении всплесков трафика во время перегрузки. Однако с учётом отсутствия корреляций в прибытии сегментов разных соединений TCP увеличение размера начального окна TCP не будет существенно повышать частоту отбрасывания сегментов. Основанное на результатах моделирования обсуждение этого вопроса приведено в параграфе 7.2.

Эти потенциальные опасности для сетей исследовались на моделях и в экспериментах, описанных ниже. На наш взгляд опасность коллапса насыщения в современной сети Internet существует (см. [FF99], где обсуждаются вопросы связанные с коллапсом насыщения в результате расширения использования соединений UDP без сквозного контроля насыщения), увеличение размера начального окна TCP до 4 Кбайт не ведёт к эскалации опасности.

6. Взаимодействие с таймером повтора передачи

Использование увеличенного начального «всплеска» данных может усугубить имеющиеся проблемы за счёт возникновения ложных тайм-аутов повтора на низкоскоростных путях в предположении использования стандартного алгоритма TCP для определения тайм-аута повтора (RTO¹) [RFC2988]. Проблема заключается в том, что на низкоскоростных путях, где время передачи составляет значительную часть периода кругового обхода, мелкие пакеты, используемые для организации соединения TCP, не обеспечивают корректной оценки RTO. При передаче первого окна данных отсчёт таймера повторной передачи на стороне отправителя может завершиться до получения подтверждений для пакетов первого окна. По прибытии таких подтверждений таймер в общем случае сбрасывается. Таким образом, при отсутствии подтверждений тайм-аут будет наступать по прошествии по крайней мере 1 секунды, в соответствии со значением RTO, рекомендуемым в RFC 2988.

В качестве примера рассмотрим канал с полосой 9,6 кбит/с. Начальное измерение RTT будет давать время кругового обхода около 67 мсек, если мы будем просто рассматривать передачу двух пакетов (SYN и SYN-ACK) размером по 40 байтов. Используя оценку RTO в соответствии с [RFC2988], мы получим начальное значение RTO = 201 мсек ($67 + 4 \cdot (67/2)$). Однако это значение RTO будет округляться до 1 секунды в соответствии с RFC 2988. Предположим, что передаётся начальное окно, содержащее один или несколько пакетов по 1500 байтов (1460 байтов данных и заголовки). Каждый пакет будет передаваться примерно 1,25 сек. Следовательно, время RTO наступит раньше, чем будет получен сегмент ACK для первого пакета, что приведёт к возникновению ложного тайм-аута. В этом случае увеличенное начальное окно размером в 3 или 4 пакета усугубит проблемы, связанные с ложным тайм-аутом.

Одним из способов решения этой проблемы является повышение консервативности алгоритма RTO. В начальном окне, например, можно обновлять значение RTO при получении каждого подтверждения. Кроме того, если отсчёт таймера повтора завершается для потерянного пакета из первого окна данных, можно задержать экспоненциальное снижение значения таймера повтора по крайней мере до одного корректного измерения RTT, включающего принятый пакет.

Другой метод заключается в отказе от измерения RTT в процессе организации соединения и сохранении принятого по умолчанию RTO до тех пор, пока измерение не будет проведено с использованием сегмента данных TCP и соответствующего сегмента ACK. Хотя этот метод позволяет предотвратить ненужные повторы передачи, он может также замедлить передачу данных в случаях, когда пакеты будут теряться до установки начального значения RTO. Использование ограниченной передачи [RFC3042] при восстановлении соединения TCP после потери с помощью ускоренного повтора вместо ожидания таймера RTO позволяет снизить негативное влияние, вызванное использованием принятого по умолчанию большого значения RTO на этапе передачи начального окна данных.

Эта реализация оставляет решение вопроса выбора действий (или бездействия) в части установки значения RTO при использовании увеличенного начального окна за разработчиками. Однако **рекомендуется** воздерживаться от

¹Retransmission timeout.

измерения RTT в процессе трехэтапного согласования с сохранением принятого по умолчанию значения RTO до тех пор, пока не будет определено значение RTT, включающее передачу пакета данных. Кроме того, **рекомендуется** использовать для TCP ограниченную передачу [RFC3042].

7. Типичные пиковые уровни для трафика TCP

Более крупные начальные окна TCP не будут значительно увеличивать всплески трафика TCP в сегодняшней сети Internet, поскольку такой трафик уже достаточно «взрывной». Всплески в 2 и 3 сегмента уже стали нормальными для TCP [Flo97]; задержавшееся подтверждение ACK (покрывающее два ранее не подтверждённых сегмента), полученное в результате предотвращения перегрузок сдвигает окно насыщения и вызывает передачу двух сегментов. Такое же задержанное подтверждение ACK, полученное во время замедленного старта, сдвигает окно на два сегмента и увеличивает его на один сегмент, что приводит к передаче трёх сегментов сразу. Всплески из 4 или 5 сегментов не столь типичны, но и не редки. В предположении задержки ACK одно отброшенное подтверждение ACK вынуждает последующее подтверждение ACK покрывать 4 ранее не подтверждённых сегмента. В процессе предотвращения перегрузки это вызывает всплеск из четырёх сегментов, а при замедленном старте - из пяти.

Готовится множество предложений, способных ослабить проблемы производительности, вызываемые умеренными всплесками трафика. Одно из таких предложений включает развёртывание более скоростных каналов в некоторых частях сети, где всплески трафика в 4 Кбайта могут представлять незначительную часть данных. Второе изменение относится к маршрутизаторам с достаточным объёмом буферов и состоит в развёртывании механизмов управления очередями типа RED, которые обеспечивают устойчивость к непродолжительным всплескам трафика.

8. Результаты моделирования и экспериментов

8.1 Исследование соединений TCP с увеличенным начальным окном

В этом параграфе рассмотрено моделирование и эксперименты по влиянию больших начальных окон на соединения TCP. В первой группе экспериментов изучалась производительность на спутниковых каналах. Большие начальные окна показали повышение производительности соединений TCP на спутниковых каналах [All97b]. В этих экспериментах начальное окно в 4 сегмента (MSS 512 байтов) повышало производительность до 30% (в зависимости от размера передачи). В [KAGT98] показано, что использование больших начальных окон снижает время передачи в тестах HTTP через спутниковые каналы ACTS. Моделирование большого числа транзакций HTTP через среду HFC¹ показывает, что увеличение начального размера окна уменьшает время загрузки страниц WWW [Nic98].

Во второй группе экспериментов изучалась работа TCP на коммутируемых модемных каналах. На соединениях 28,8 кбит/с² [All97a, AHO98] 4-сегментное начальное окно снижало время передачи файла размером 16 Кбайт примерно на 10% без роста частоты потери пакетов. В работе [RFC2416] исследовалось влияние использования увеличенного начального окна при подключении хоста по медленному модемному каналу к маршрутизатору с буфером на 3 пакета. По результатам исследования отмечено, что увеличение начального размера окна не оказывает негативного влияния на производительность TCP.

В работе [All00] показано, что доля соединений на конкретном web-сервере, сталкивающихся с потерей пакетов в начальном окне передачи данных, возрастает с увеличением размера начального окна насыщения. Однако это увеличение соответствует потерям, ожидаемым в результате передачи в сеть существенного объёма трафика.

8.2 Изучение сетей, использующих увеличенный размер начального окна

В этом параграфе рассматриваются экспериментальные и модельные исследования влияния большого окна TCP на другие соединения в том же пути. Эксперименты [All97a, AHO98] показали, что передачи по 16 Кбайт на 100 хостов Internet с 4-сегментным начальным окном приводят к незначительному росту частоты отбрасывания пакетов (0,04 сегмента на передачу). Частота потерь возросла незначительно, а время передачи уменьшалось приблизительно на 25% при использовании начального окна размером 4 сегмента (MSS 512 байт) по сравнению с окном в 1 сегмент.

Моделирование в [RFC2415] было связано с изучением воздействия увеличения начального окна на одновременный сетевой трафик. В этой работе потоки HTTP и FTP использовали общий перегруженный шлюз (число потоков HTTP и FTP варьировалось от случая к случаю). Для каждого случая определялась общая загрузка канала и частота отбрасывания пакетов, средняя задержка при загрузке страницы web и скорость передачи FTP. Увеличение начального окна обычно приводило к росту пропускной способности, незначительному росту уровня потери пакетов и повышению общей производительности сети. За исключением одного случая увеличение начального окна повышало частоту потерь не более, чем на 1% по сравнению с 1-сегментным начальным окном - потери достигали 3,5% при односегментном окне и 4,5% - при 4-сегментном. Вывод этой работы заключался в том, что увеличение начального окна TCP до трёх пакетов (4380 байтов) повышает производительность сети.

Morris [Mor97] исследовал влияние увеличенных начальных окон в тяжело нагруженной сети для передач размером 20 Кбайт. Частота потерь в сетях, где все соединения TCP использовали начальное окно в 4 сегмента, была на 1-2% выше чем для 1-сегментных окон. Эта зависимость сохранялась для случаев, когда потери при 1-сегментном начальном окне составляли от 1% до 11%. Кроме того, в сетях с начальным окном в 4 сегмента отправители TCP дольше ждали тайм-аута RTO для повтора передачи по сравнению с 1-сегментным окном. Это время ожидания является простым, когда через соединение не передаётся данных. Результаты показывают, что в сильно загруженной среде, где все соединения используют общий узкий канал, увеличение начального окна может приводить к заметному росту числа потерь и тайм-аутов повтора.

9. Вопросы безопасности

В этом документе обсуждается начальный размер окна насыщения, разрешенный для соединений TCP. Изменение этого размера не меняет состояния безопасности TCP.

¹Hybrid fiber coax.

²В оригинале ошибочно указана скорость 28,8 бит/с (см. http://www.rfc-editor.org/errata_search.php?rfc=3390&eid=4569). Прим. перев.

10. Заключение

Этот документ задаёт незначительное изменение TCP, которое будет очевидно полезным для краткосрочных соединений TCP, а также для соединений через каналы с большим значением RTT (обеспечивается экономия времени в несколько RTT на этапе замедленного старта).

11. Благодарности

Мы благодарим Vern Paxson, Tim Shepard, участников списка рассылки End-to-End-Interest и членов рабочей группы IETF TCP Implementation за обсуждение проблемы и отклики на этот документ.

12. Литература

- [AHO98] Mark Allman, Chris Hayes, and Shawn Ostermann, An Evaluation of TCP with Larger Initial Windows, March 1998. ACM Computer Communication Review, 28(3), July 1998. "<http://roland.lerc.nasa.gov/~mallman/papers/initwin.ps>".
- [All97a] Mark Allman. An Evaluation of TCP with Larger Initial Windows. 40th IETF Meeting - TCP Implementations WG. December, 1997. Washington, DC.
- [All97b] Mark Allman. Improving TCP Performance Over Satellite Channels. Master's thesis, Ohio University, June 1997.
- [All00] Mark Allman. A Web Server's View of the Transport Layer. ACM Computer Communication Review, 30(5), October 2000.
- [FF96] Fall, K., and Floyd, S., Simulation-based Comparisons of Tahoe, Reno, and SACK TCP. Computer Communication Review, 26(3), July 1996.
- [FF99] Sally Floyd, Kevin Fall. Promoting the Use of End-to-End Congestion Control in the Internet. IEEE/ACM Transactions on Networking, August 1999. URL "<http://www.icir.org/floyd/end2end-paper.html>".
- [FJ93] Floyd, S., and Jacobson, V., Random Early Detection gateways for Congestion Avoidance. IEEE/ACM Transactions on Networking, V.1 N.4, August 1993, p. 397-413.
- [Flo94] Floyd, S., TCP and Explicit Congestion Notification. Computer Communication Review, 24(5):10-23, October 1994.
- [Flo96] Floyd, S., Issues of TCP with SACK. Technical report, January 1996. Available from <http://www.nrg.ee.lbl.gov/floyd/>.
- [Flo97] Floyd, S., Increasing TCP's Initial Window. Viewgraphs, 40th IETF Meeting - TCP Implementations WG. December, 1997. URL "<ftp://ftp.ee.lbl.gov/talks/sf-tcp-ietf97.ps>".
- [KAGT98] Hans Kruse, Mark Allman, Jim Griner, Diepchi Tran. HTTP Page Transfer Rates Over Geo-Stationary Satellite Links. March 1998. Proceedings of the Sixth International Conference on Telecommunication Systems. URL "<http://roland.lerc.nasa.gov/~mallman/papers/nash98.ps>".
- [Mor97] Robert Morris. Частное сообщение, 1997. Указано только с целью благодарности.
- [Nic98] Kathleen Nichols. Improving Network Simulation With Feedback, Proceedings of LCN 98, October 1998. URL "<http://www.computer.org/proceedings/lcn/8810/8810toc.htm>".
- [Pos82] Postel, J., "Simple Mail Transfer Protocol", STD 10, [RFC 821](#), August 1982.
- [RFC1122] Braden, R., "Requirements for Internet Hosts -- Communication Layers", STD 3, [RFC 1122](#), October 1989.
- [RFC1191] Mogul, J. and S. Deering, "Path MTU Discovery", [RFC 1191](#), November 1990.
- [RFC1945] Berners-Lee, T., Fielding, R. and H. Nielsen, "Hypertext Transfer Protocol -- HTTP/1.0", RFC 1945, May 1996.
- [RFC2068] Fielding, R., Mogul, J., Gettys, J., Frystyk, H. and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1", RFC 2616, January 1997.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, [RFC 2119](#), March 1997.
- [RFC2309] Braden, B., Clark, D., Crowcroft, J., Davie, B., Deering, S., Estrin, D., Floyd, S., Jacobson, V., Minshall, G., Partridge, C., Peterson, L., Ramakrishnan, K., Shenker, S., Wroclawski, J. and L. Zhang, "Recommendations on Queue Management and Congestion Avoidance in the Internet", RFC 2309, April 1998.
- [RFC2414] Allman, M., Floyd, S. and C. Partridge, "Increasing TCP's Initial Window", [RFC 2414](#), September 1998.
- [RFC2415] Poduri, K. and K. Nichols, "Simulation Studies of Increased Initial TCP Window Size", RFC 2415, September 1998.
- [RFC2416] Shepard, T. and C. Partridge, "When TCP Starts Up With Four Packets Into Only Three Buffers", RFC 2416, September 1998.
- [RFC2581] Allman, M., Paxson, V. and W. Stevens, "TCP Congestion Control", [RFC 2581](#), April 1999.
- [RFC2821] Klensin, J., "Simple Mail Transfer Protocol", [RFC 2821](#), April 2001.
- [RFC2988] Paxson, V. and M. Allman, "Computing TCP's Retransmission Timer", RFC 2988, November 2000.
- [RFC3042] Allman, M., Balakrishnan, H. and S. Floyd, "Enhancing TCP's Loss Recovery Using Limited Transmit", [RFC 3042](#), January 2001.
- [RFC3168] Ramakrishnan, K.K., Floyd, S. and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", [RFC 3168](#), September 2001.

¹Статья по указанной в тесте ссылке недоступна. Можно воспользоваться другой [ссылкой](#). Прим. перев.

Приложение А - Дубликаты сегментов

В современных средах (без ECN¹ [Flo94] [RFC2481]) все TCP используют факты отбрасывания сегментов, как индикацию от сети о достижении предела пропускной способности. В этом документе утверждается, что увеличение начального размера окна не должно приводить к росту числа передаваемых отправителем повторов для пакетов, которые уже приняты получателем.

Если из начального окна отбрасывается один сегмент, восстановление TCP может пойти тремя путями: (1) замедленный старт с начальным окном в один сегмент, как после тайм-аута повтора или ускоренного повтора в Tahoe TCP; (2) быстрое восстановление без селективных подтверждений (SACK²), как после трёх дубликатов ACK в Reno TCP; (3) быстрое восстановление с SACK для TCP, где отправитель и получатель поддерживают опцию SACK [MMFR96]. Во всех трёх случаях отбрасывание сегмента из начального окна не приводит к передаче дубликатов (т. е., сегментов, которые уже были приняты получателем). Отметим, что для TCP, передающего 4 сегмента по 512 байтов в начальном окне, отбрасывание одного сегмента не будет требовать тайм-аута повтора и восстановление может быть выполнено с помощью алгоритма Fast Retransmit (если отсчёт таймера повтора не завершился раньше). При отбрасывании одного сегмента из начального окна в 3 сегмента восстановление возможно с помощью алгоритма ускоренного повтора в зависимости от того, какой из сегментов был отброшен и используются ли незадержанные подтверждения ACK. Например, отбрасывание первого из трёх сегментов начального окна всегда будет требовать ожидания тайм-аута в отсутствие Limited Transmit [RFC3042]. Однако отбрасывание третьего сегмента всегда позволяет восстановление с помощью алгоритма ускоренного повтора, если подтверждения ACK не терялись.

Рассмотрим случай, когда начальное окно имеет размер от 2 до 4 сегментов и по крайней мере два из этих сегментов отбрасываются. Если отброшены все сегменты начального окна, ясно, что возникновение дубликатов просто невозможно, поскольку получатель ещё не принял ничего (вопрос расхода дефицитной полосы, связанного с отброшенными сегментами, был рассмотрен в разделе 5).

При отбрасывании двух сегментов из 3-сегментного начального окна отправитель будет передавать дубликат сегмента только в том случае, когда отброшены два первых сегмента из трёх и отправителем не было получено пакета с опцией SACK, подтверждающей третий сегмент.

При отбрасывании двух сегментов из 4-сегментного начального окна проверка 6 возможных вариантов (не проводится здесь) показывает, что в зависимости от номеров отброшенных пакетов в отсутствие SACK отправитель может передать один сегмент-дубликат. Вариантов с передачей двух дубликатов не возникает.

При отбрасывании трёх сегментов из 4-сегментного начального окна в отсутствие SACK возможна отправка одного дубликата (в зависимости от номеров отброшенных сегментов).

Таким образом, при отсутствии SACK и отбрасывании множества сегментов из начального окна возможны ситуации с передачей одного сегмента-дубликата. Передачи нескольких дубликатов не возникает ни в одном случае. В результате можно сделать вывод о том, что вероятность появления на приёмной стороне сегментов-дубликатов в результате увеличения начального окна должна быть достаточно мала.

Адреса авторов

Mark Allman

BBN Technologies/NASA Glenn Research Center
21000 Brookpark Rd
MS 54-5
Cleveland, OH 44135
E-Mail: mallman@bbn.com
<http://roland.lerc.nasa.gov/~mallman/>

Sally Floyd

ICSI Center for Internet Research
1947 Center St, Suite 600
Berkeley, CA 94704
Phone: +1 (510) 666-2989
E-Mail: floyd@icir.org
<http://www.icir.org/floyd/>

Craig Partridge

BBN Technologies
10 Moulton St
Cambridge, MA 02138
E-Mail: craig@bbn.com

Перевод на русский язык

Николай Малых

nmalykh@protokols.ru

Полное заявление авторских прав

Copyright (C) The Internet Society (2002). Все права защищены.

Этот документ и его переводы могут копироваться и предоставляться другим лицам, а производные работы, комментирующие или иначе разъясняющие документ или помогающие в его реализации, могут подготавливаться, копироваться, публиковаться и распространяться целиком или частично без каких-либо ограничений при условии

¹Explicit Congestion Notification - явное уведомление о насыщении.

²Selective acknowledgment.

сохранения указанного выше уведомления об авторских правах и этого параграфа в копии или производной работе. Однако сам документ не может быть изменён каким-либо способом, таким как удаление уведомления об авторских правах или ссылок на Internet Society или иные организации Internet, за исключением случаев, когда это необходимо для разработки стандартов Internet (в этом случае нужно следовать процедурам для авторских прав, заданных процессом Internet Standards), а также при переводе документа на другие языки.

Предоставленные выше ограниченные права являются бессрочными и не могут быть отозваны Internet Society или правопреемниками.

Этот документ и содержащаяся в нем информация представлены "как есть" и автор, организация, которую он/она представляет или которая выступает спонсором (если таковой имеется), Internet Society и IETF отказываются от каких-либо гарантий (явных или подразумеваемых), включая (но не ограничиваясь) любые гарантии того, что использование представленной здесь информации не будет нарушать чьих-либо прав, и любые предполагаемые гарантии коммерческого использования или применимости для тех или иных задач.

Подтверждение

Финансирование функций RFC Editor обеспечено Internet Society.