Network Working Group
Request for Comments: 2481
Category: Experimental

K. Ramakrishnan
AT&T Labs Research
S. Floyd
LBNL

January 1999

Явное уведомление о насыщении (ECN)

A Proposal to add Explicit Congestion Notification (ECN) to IP

Статус документа

В этом документе определен экспериментальный протокол, предлагаемый сообществу Internet. Документ не содержит каких-либо стандартов Internet. Документ служит приглашением к дискуссии с целях совершенствования протокола и может распространяться без ограничений.

Авторские права

Тезисы

В данном документе описывается предлагаемое добавление флага ECN[1] в IP. Протокол TCP в настоящее время является доминирующим протоколом транспортного уровня в Internet. Начнем с описания процедуры отбрасывания пакетов протоколом TCP в качестве индикации насыщения (перегрузки). После этого будет показано, что добавление активного управления очередями (например, RED[2]) в инфраструктуру Internet, в результате которого маршрутизаторы смогут детектировать насыщение до момента переполнения очереди, позволит маршрутизаторам избавиться от необходимости отбрасывания пакетов для индикации перегрузки. Маршрутизаторы смогут вместо отбрасывания устанавливать флаг CE[3] в заголовке пакетов поддерживающих ECN транспортных протоколов. Описывается процедура установки флага CE в маршрутизаторах, а также изменения, которые нужно внести в протокол TCP для поддержки ECN. Изменения других протоколов транспортного уровня (например, негарантированная доставка пакетов unicast или multicast, гарантированная доставка multicast-пакетов, другие протоколы гарантированной доставки пакетов unicast) могут рассматриваться при разработке или стандартизации таких протоколов.

1. Соглашение об использовании терминов

Ключевые слова **необходимо** (MUST), **недопустимо** (MUST NOT), **требуется** (REQUIRED), **нужно** (SHALL), **не нужно** (SHALL NOT), **следует** (SHOULD), **не следует** (SHOULD NOT), **рекомендуется** (RECOMMENDED), **возможно** (MAY), **необязательно** (OPTIONAL) в данном документе должны интерпретироваться в соответствии с [B97].

2. Введение

Алгоритмы контроля насыщения и предотвращения перегрузки протокола TCP основаны на представлении сети, как «черного ящика» [Jacobson88, Jacobson90]. Насыщение или его отсутствие определяется конечными системами путем проверки состояния сети за счет увеличения нагрузки (расширения окна насыщения - числа пакетов, остающихся в сети) до тех пор, пока не возникнет насыщение и связанная с ним потеря пакетов. Трактовка сети, как «черного ящика», и потери пакетов, как индикатора перегрузки, приемлема для случаев передачи по протоколу TCP данных, которые не чувствительны или не критичны к задержкам или потере отдельных пакетов. Алгоритмы контроля насыщения в TCP используют встроенные методы (такие, как Fast Retransmit и Fast Recovery[4]) для минимизации влияния потерь с точки зрения пропускной способности.

Однако эти алгоритмы не подходят для приложений, которые чувствительны к задержкам или потере одного или нескольких пакетов. Интерактивные системы (например, telnet, просмотр web-страниц, передача аудио/видео-потоков) могут быть чувствительны к потере пакетов (при использовании транспорта без гарантии доставки типа UDP) или увеличению задержки, вызванному повтором передачи в результате потери пакета (при использовании гарантированного транспорта типа TCP).

Поскольку TCP определяет приемлемый размер окна насыщения путем увеличения размера этого окна до тех пор, пока не начнется потеря (отбрасывание) пакетов, это может приводить к переполнению очередей в маршрутизаторах, являющихся узким местом в сети. Большинство алгоритмов отбрасывания пакетов в маршрутизаторах не чувствительно к нагрузке, создаваемой отдельным потоком, что означает возможность отбрасывания пакетов из некоторых потоков, чувствительных к задержкам. Активные механизмы управления очередями детектируют насыщение до того, как переполнится очередь, и обеспечивают индикацию насыщения для конечных узлов. Преимущества активного управления очередями обсуждаются в RFC 2309 [RFC2309]. Такое управление позволяет избавиться от некоторых негативных свойств систем управления очередями, основанных на отбрасывании пакетов при переполнении (в частности, от нежелательной синхронизации потери пакетов во множестве потоков данных). Более важно, что в системах активного управления очередями транспортные протоколы с контролем насыщения (например, TCP) не используют переполнение буферов в качестве индикатора перегрузки. Это может снизить избыточные задержки в очереди для всех типов трафика, использующих эту очередь.

---

[1] Explicit Congestion Notification - явное уведомление о насыщении.
[2] Random Early Detection - упреждающее детектирование.
[3] Congestion Experienced - наблюдается насыщение.
[4] Ускоренный повтор передачи и быстрое восстановление

Механизмы активного управления очередями могут использовать один из нескольких методов индикации насыщения для конечных узлов. Одним из методов является отбрасывание пакетов, как это делается сейчас. Однако активное управление очередями позволяет маршрутизатору отделить правила буферизации (включения в очередь) и отбрасывания пакетов от политики индикации насыщения. Таким образом, активное управление очередями позволяет маршрутизаторам использовать бит CE в заголовке пакета для индикации насыщения вместо того, чтобы полагаться исключительно на отбрасывание пакетов.

3. Допущения и общие принципы

В этом параграфе рассматриваются некоторые важные принципы и допущения, на которых основана работа предлагаемого расширения.

(1) Длительность периода насыщения может меняться в широких пределах. Периоды насыщения могут превышать время кругового обхода (RTT).

(2) Число пакетов в отдельном потоке (например, в соединении TCP или сеансе обмена данными по протоколу UDP) также может меняться в широких пределах. Мы заинтересованы в контроле насыщения, создаваемого потоком, который передает достаточно большое количество данных, чтобы такой поток оставался активным до того, как будет получен сигнал обратной связи из сети.

(3) Новые механизмы контроля и предотвращения перегрузки должны сосуществовать и кооперироваться с существующими механизмами контроля насыщения. В частности, новые механизмы должны сосуществовать с методами, используемыми в TCP, и принятой в современных маршрутизаторах практикой отбрасывания пакетов в периоды насыщения.

(4) Очевидно, что адаптация ECN займет достаточно продолжительное время, поэтому важное значение приобретает процесс перехода. Некоторые маршрутизаторы могут по-прежнему использовать для индикации насыщения лишь отбрасывание пакетов, а некоторые конечные системы могут не поддерживать ECN. Наиболее жизненным будет процесс перехода, при котором не будет происходить деления сети на «островки», поддерживающие и не поддерживающие ECN.

(5) Очевидно, что асимметрия маршрутов является нормальным явлением в Internet. Путь (последовательность каналов и маршрутизаторов), по которому данные следуют из одной точки в другую в одном направлении, может отличаться от пути между той же парой точек в обратном направлении (например, для передачи подтверждений).

(6) Многие маршрутизаторы более эффективно обрабатывают заголовки пакетов IP без опций, нежели опции IP. Этот факт служит предпосылкой включения индикации насыщения в обычный заголовок пакета IP, а не в поле опций заголовка.

(7) Следует признать, что не все конечные системы могут кооперироваться в плане контроля насыщения. Однако новым механизмам не следует упрощать для приложений TCP запрет контроля насыщения. Преимущества от использования новых механизмов типа ECN не должны быть значительными.

4. Упреждающее детектирование перегрузки

Упреждающее детектирование RED представляет собой механизм активного управления очередями, предложенный для детектирования начинающейся перегрузки [FJ93] и развернутый уже в магистральных сетях Internet [RFC2309]. Хотя RED предложен в качестве механизма общего назначения для индикации насыщения, в современной среде Internet использование RED ограничено отбрасыванием пакетов для индикации насыщения. RED отбрасывает пакеты в тех случаях, когда средний размер очереди превышает пороговое значение, не дожидаясь переполнения очереди. Однако, когда RED отбрасывает пакеты до переполнения очереди, это отбрасывание не вызвано нехваткой памяти.

RED может устанавливать бит CE в заголовке пакета вместо того, чтобы отбросить пакет, если такой бит присутствует в заголовке IP и понятен транспортному протоколу. Использование бита CE будет позволять транспортному протоколу получателя избавиться от избыточной задержки, связанной с повтором передачи после отбрасывания пакета. Для обозначения пакетов с установленным битом CE далее будет использоваться термин CE-пакет.

5. Явное уведомление о насыщении в IP

Предлагается обеспечивать в Internet индикацию насыщения для наступающей перегрузки (как в RED и более ранней работе [RJ90]), когда уведомление о возможной перегрузке, выполняемое путем маркировки пакетов, может произойти раньше, нежели начнется отбрасывание пакетов. Для использования этого метода требуется добавить двухбитовое поле ECN в заголовок IP. Бит ECT[1] будет устанавливаться отправителем данных для индикации удаленной точке поддержки ECN на уровне транспортного протокола. Бит CE будет устанавливаться маршрутизатором для индикации насыщения конечным узлам. Маршрутизаторы, получающие пакеты при заполненных очередях, будут отбрасывать такие пакеты, как это делалось и раньше.

Для поля ECN предлагается использовать биты 6 и 7 октета TOS в заголовках IPv4. Бит 6 используется для флага ECT, а бит 7 – для CE. Октету TOS в заголовке IPv4 соответствует октет Traffic Class в заголовке IPv6. Определения для октетов IPv4 TOS [RFC791] и IPv6 Traffic Class должны быть заменены полем DS (Differentiated Services) [DIFFSERV]. Биты 6 и 7 указаны в [DIFFSERV], как неиспользуемые (Currently Unused). В параграфе 19 приведена краткая история использования октета TOS.

По причине изменений в характере использования TOS описанное здесь применение поля ECN не может гарантировать совместимости со всеми прежними вариантами использования этих двух битов. Потенциальные проблемы, связанные с отсутствием совместимости, рассматриваются в параграфе 19.

При получении поддерживающим ECN транспортным протоколом одного CE-пакета алгоритм контроля насыщения в конечной системе **должен** работать в точности так же, как при отбрасывании одного пакета. Например, для поддерживающей ECN реализации протокола TCP требуется, чтобы отправитель уменьшил вдвое размер окна насыщения в ответ на любой факт отбрасывания пакета или получения индикатора ECN. Однако следует отметить, что существуют некоторые существенные различия в реакции отправителя TCP, связанные деталями поведения протокола

[1] ECN-Capable Transport – поддерживающий ECN транспорт.

в ответ на индикацию насыщения. Для реакции TCP на получение индикации ECN не рекомендуется выполнение таких процедур, как замедленный старт (slow-start в Tahoe TCP) в ответ на отбрасывание пакетов или ожидание Reno TCP в течение периода около половины времени кругового обхода при использовании быстрого восстановления (Fast Recovery).

Одной из причин требования идентичной реакции на индикацию насыщения при получении CE-пакета или отбрасывании пакета является необходимость обеспечения постепенного развертывания ECN как в конечных системах, так и в маршрутизаторах. Некоторые маршрутизаторы могут отбрасывать ECN-пакеты[1] (например, при использовании некоторых правил детектирования перегрузки в RED), тогда как другие маршрутизаторы будут устанавливать бит CE при таких же условиях насыщения. Подобно этому, маршрутизатор может отбрасывать не поддерживающие ECN пакеты или устанавливать бит CE в ECN-пакетах при одинаковых условиях насыщения. Разная реакция на индикацию насыщения путем установки бита CE и путем отбрасывания пакетов может приводить к различным (неправильным) трактовкам для разных потоков.

Дополнительное требование заключается в том, что конечным системам следует реагировать на насыщение не более одного раза в расчете на окно данных (т. е., не более одного раза за период кругового обхода), чтобы избежать множественной реакции на несколько фактов индикации насыщения в течение одного периода кругового обхода.

Маршрутизаторам следует устанавливать бит CE в ECN-пакетах только в тех случаях, когда маршрутизатор должен был бы отбросить пакет для индикации насыщения конечным узлам. Когда буфер маршрутизатора еще не заполнен, но маршрутизатор уже приготовился к отбрасыванию пакетов для уведомления конечных узлов о насыщении, этому маршрутизатору следует сначала проверить наличие бита ECT в заголовке IP. Если данный бит установлен, вместо отбрасывания пакетов маршрутизатор **может** устанавливать бит CE в заголовке IP.

Среда, где все конечные узлы поддерживают ECN, позволяет разработать новые критерии установки бита CE и новые механизмы контроля насыщения для реакции конечных узлов на CE-пакеты. Однако это является темой для исследования и выходит за пределы данного документа.

Когда маршрутизатор получает CE-пакет, бит CE остается без изменения и пакет передается как обычно. При возникновении некоторой перегрузки и заполнении очереди маршрутизатора, последний может принять решение об отбрасывании некоторых пакетов при поступлении новых. Предполагается, что такое отбрасывание пакетов станет сравнительно редким явлением, когда большая часть конечных систем будет поддерживать ECN и механизмы контроля насыщения TCP или аналогичные механизмы. В корректно организованной сети, работающей в среде с поддержкой ECN, потери пакетов могут происходить в периоды неустойчивости или в присутствии не поддерживающих контроль насыщения отправителей.

Предполагается, что маршрутизаторы будут устанавливать бит CE в ответ на начинающуюся перегрузку, указываемую средним размером очереди, с использованием алгоритма RED, предложенного в [FJ93, RFC2309]. По имеющимся у авторов сведениям этот вариант является единственным предложением, обсуждаемым IETF, для упреждающего отбрасывания пакетов маршрутизаторами до переполнения буферов. Однако данный документ не пытается задавать тот или иной механизм активного управления очередями, оставляя решение этого вопроса (если он возникнет) за IETF. Хотя использование ECN связано с вопросом о необходимости иметь подходящий механизм активного управления очередями в маршрутизаторах, авторы не настаивают на использовании именно этого механизма. Методы активного управления очередями были разработаны и развернуты независимо от ECN с отбрасыванием пакетов для индикации насыщения еще до начала использования ECN в архитектуре IP.

6. Поддержка со стороны транспортного протокола

ECN требует поддержки со стороны транспортного протокола в дополнение к функциональности, обеспечиваемой полем ECN в заголовке пакета IP. Транспортный протокол может требовать согласования между конечными точками при организации соединения для проверки поддержки ими ECN, чтобы отправитель мог устанавливать код ECT в передаваемых пакетах. Во-вторых, транспортный протокол должен быть способен соответствующим образом реагировать на получение пакетов CE. Эта реакция может выражаться в форме информирования получателем отправителя данных о полученном пакете CE (например, TCP), отказа получателя от участия в многоуровневой multicast-группе (например, RLM [MJV96]) или в ином виде, обеспечивающем, в конечном итоге, снижение скорости поступления потока данных для этого получателя.

В этом документе рассматривается добавление поддержки ECN[2] только для протокола TCP, а рассмотрение вопросов использования ECN в других транспортных протоколах оставлено для будущих исследований. Для TCP добавление ECN требует поддержки трех новых механизмов - согласование между конечными точками в процессе организации соединения для проверки поддержки ECN на обеих сторонах; флаг ECN-Echo (ECE) в заголовке TCP, чтобы получатель мог информировать отправителя о получении пакета CE; флаг CWR[3] в заголовке TCP, чтобы отправитель мог информировать получателя о снижении размера окна насыщения. Очевидно, что функциональность, требуемая от других протоколов (в частности, протоколов групповой адресации с гарантиями доставки и без таковых), будет отличаться и определится при стандартизации этих транспортных протоколов в IETF.

В этом документе используется термин «пакеты TCP» вместо «сегментов TCP», что не вполне корректно терминологически.

6.1. TCP

В последующих параграфах детально рассматривается предложенное использование ECN в TCP. Эти предложения представлены в той же форме, что и в работе [Floyd94]. Предполагается, что TCP на стороне отправителя использует стандартные алгоритмы контроля насыщения Slow-start, Fast Retransmit и Fast Recovery [RFC2001].

Это предложение задает два новых флага в резервном поле заголовка TCP. Механизм TCP для согласования поддержки ECN использует флаг ECE[4] (ECN-Echo) в заголовке TCP. Бит 9 в поле Reserved заголовка TCP

[1] В оригинале “ECN-Capable” - пакеты, для которых может поддерживаться ECN. *Прим. перев.*
[2] ECN Capability.
[3] Congestion Window Reduced - окно насыщения уменьшено.
[4] В более ранних документах этот флаг назывался ECN Notify.

предназначен для использования в качестве флага ECN-Echo. Местоположение шестибитового резервного поля заголовка TCP показано на рисунке 3 в RFC 793 [RFC793].

Чтобы обеспечить получателю TCP возможность определения момента прекращения установки флага ECN-Echo, в заголовок TCP добавлен еще один флаг - CWR. Для флага CWR выделен бит 8 резервного поля в заголовке TCP .

Использование этих флагов описано ниже.

6.1.1. Инициализация TCP

На этапе организации соединения TCP модули TCP на стороне отправителя и получателя обмениваются информацией о своем намерении использовать ECN. После завершения этого согласования отправитель TCP устанавливает код ECT в заголовке IP пакета данных для индикации сетевым устройствам возможности и желания использовать ECN для этого пакета. Этот код показывает маршрутизаторам, что они могут маркировать данный пакет кодом CE, если они хотят использовать такой метод индикации насыщения. Если соединение TCP не хочет использовать ECN-уведомление для отдельного пакета, передающая сторона TCP устанавливает в качестве кода ECN значение 0 (т. е., не устанавливает флаг), а получатель TCP игнорирует код CE в полученном пакете.

Когда узел передает пакет TCP SYN, он может установить в заголовке TCP флаги ECN-Echo и CWR. Для пакетов SYN установка флагов ECN-Echo и CWR определена, как индикация того, что передающая сторона TCP поддерживает ECN, а не для индикации насыщения или отклика на насыщение. Точнее говоря, пакет SYN с установленными флагами ECN-Echo и CWR показывает, что передающая пакет SYN реализация TCP будет участововать в ECN в качестве отправителя и получателя. В качестве получателя она будет отвечать на входящие пакеты с флагом CE в заголовке IP установкой флага ECN-Echo в исходящих пакетах TCP ACK. В качестве отправителя она будет отвечать на входящие пакеты с флагом ECN-Echo снижением размера окна насыщения, когда это приемлемо.

Когда хост узел передает пакет SYN-ACK, он может установить флаг ECN-Echo, но не устанавливать флаг CWR. Для пакетов SYN-ACK с установленным флагом ECN-Echo и сброшенным флагом CWR в заголовке TCP опрпределяется, как индикация того, что узел TCP, передавший пакет SYN-ACK, поддерживает ECN.

Возникает вопрос, почему для пакетов SYN используется два связанных с ECN флага в резервном поле заголовка TCP, тогда как в ответном пакете SYN-ACK устанавливается только один связанный с ECN флаг. Такая асимметрия нужна для отказоустойчивого согласования поддержки ECN с некоторыми имеющимися реализациями TCP. Существует по крайней мере одна некорректно работающая реализация TCP, в которой получатели устанавливают в поле Reserved в заголовке TCP пакетов ACK (и, следовательно, SYN-ACK) просто отражение поля Reserved из заголовка TCP принятого пакета данных. Поскольку в пакете TCP SYN для индикации поддержки ECN устанавливаются флаги ECN-Echo и CWR, а в пакетах SYN-ACK - только флаг ECN-Echo, передающая сторона TCP корректно интерпретирует отражение получателем своих флагов, как индикацию отсутствия поддержки ECN на приемной стороне.

6.1.2. Отправитель TCP

Для соединения TCP, использующего ECN, пакеты данных передаются с установленным (1) флагом ECT в заголовке IP. Если отправитель получает пакет ECN-Echo ACK (т. е., пакет ACK с флагом ECN-Echo в заголовке TCP), это означает, что на пути между отправителем и получателем имеется насыщение. Индикацию насыщения следует трактовать просто, как потерю в результате насыщения для TCP без поддержки ECN. Т. е., отправитель TCP снижает вдвое размер окна насыщения cwnd и уменьшает порог замедленного старта ssthresh. Передающему модулю TCP **не следует** увеличивать окно насыщения в ответ на получение пакета ECN-Echo ACK.

Критическим условием является то, что TCP не реагирует на индикацию насыщения более одного раза в каждом окне данных (или более одного раза за период кругового обхода). Т. е., окно насыщения у отправителя TCP следует уменьшать однократно в ответ на серию отброшенных или помеченных CE пакетов из одного окна данных. Кроме того, отправителю TCP не следует уменьшать значение порога ssthresh, если оно уже было снижено в последний период кругового обхода. Однако отбрасывание повторно переданных пакетов интерпретируется отправителем TCP, как новый факт перегрузки.

После того, как отправитель TCP уменьшает окно насыщения в ответ на пакет CE, входящие подтверждения, которые продолжают приходить, могут влиять на передачу исходящих пакетов, дозволенных уменьшенным окном насыщения. Если окно насыщения содержит только 1 MSS (максимальный размер сегмента) и передающий модуль TCP получает пакет ECN-Echo ACK, передающему TCP следует, в принципе, продолжать уменьшение окна насыщения вдвое. Однако размер окна насыщения ограничен снизу значением в 1 MSS. Если передающий модуль TCP будет продолжать передачу, используя окно насыщения размером 1 MSS, это приведет к передаче одного пакета за период кругового обхода. Мы полагаем, что желательно дальнейшее снижение скорости передачи TCP в ответ на получение пакета ECN-Echo при окне насыщения размером 1 MSS. Мы используем таймер повтора передачи в качестве меры снижения скорости в таких ситуациях. Поэтому, передающему модулю TCP следует сбрасывать таймер повтора при получении пакета ECN-Echo в случае единичного размера окна насыщения. Передающий модуль TCP в результате сможет передать новый пакет только после завершения отсчета таймера повтора.

В работе [Floyd94] обсуждается реакция TCP на ECN более детально. В работе [Floyd98] рассматривается тест с использованием эмулятора ns, который иллюстрирует множество сценариев ECN, включая ECN, за которым следует другой ECN, Fast Retransmit или Retransmit Timeout, Retransmit Timeout или Fast Retransmit, за которым следует ECN, окно насыщения в один пакет, за которым следует ECN.

TCP следует существующим алгоритмам передачи пакетов данных в ответ на прием пакетов ACK, дубликаты подтверждений или тайм-аут повтора [RFC20081].

6.1.3. Получатель TCP

Когда TCP принимает пакет данных CE на стороне получателя, приемный модуль TCP устанавливает флаг ECN-Echo в заголовке TCP следующего пакета ACK. Если на приемной стороне уже есть ожидающий пакет ACK (как в современных реализациях TCP с задержкой подтверждений, передающих пакет ACK по прибытии каждого второго пакета данных), тогда флаг ECN-Echo устанавливается в пакете ACK, если код CE был установлен для любого из подтверждаемых пакетов данных. Т. е., если в любом из подтверждаемых пакетов имеется маркировка CE, возвращаемый пакет ACK будет иметь флаг ECN-Echo.

Для обеспечения устойчивости к отбрасыванию пакетов ACK с флагом ECN-Echo, получатель TCP устанавливает этот флаг в передаваемых впоследствии пакетах ACK. Прекращение передачи флага ECN-Echo получатель TCP инициирует при получении флага CWR в пакете данных от передающей стороны TCP.

Когда поддерживающий ECN модуль TCP снижает размер окна насыщения по любой причине (в результате тайм-аута, Fast Retransmit или в ответ на уведомление ECN), TCP устанавливает флаг CWR в заголовке TCP первого пакета данных после снижения размера окна. Если такой пакет данных отбрасывается в сети, передающая сторона TCP будет снова снижать размер окна насыщения и заново передавать отброшенный пакет. Таким образом, сообщение о снижении размера окна насыщения (CWR) гарантированно доставляется получатель данных.

После того, как получатель TCP передает пакет ACK с установленным флагом ECN-Echo, он продолжает устанавливать этот флаг во всех передаваемых пакетах ACK (подтверждающих как пакеты данных с маркером CE, так и пакеты данных без маркера), пока не получит пакет с флагом CWR. После получения пакета CWR подтверждения для последующих пакетов без маркера CE передаются без флага ECN-Echo. Если получатель данных принимает другой пакет CE, он снова начинает передавать пакеты ACK с флагом ECN-Echo. Хотя получение пакета CWR не гарантирует получение отправителем пакета с флагом ECN-Echo, это событие говорит о том, что отправитель уменьшил размер окна насыщения в какой-то момент «после» того, как он передал пакет данных, для которого был установлен маркер CE.

Выше уже было отмечено, что отправитель TCP не должен снижать размер окна насыщения более одного раза в окне данных. Требуются некоторые меры по предотвращению многократного снижения размера окна, когда окно данных включает как отброшенные пакеты, так и пакеты с маркером CE. Этот вопрос рассматривается в работе [Floyd98].

6.1.4. Насыщение на пути ACK

В существующих реализациях механизмов контроля насыщения TCP чистые пакеты подтверждения (т. е., пакеты, содержащие только подтверждение без дополнительных данных) **должны** передаваться с кодом not-ECT. Современные получатели TCP не имеют механизмов снижения трафика на пути пакетов ACK в ответ на индикацию насыщения. Механизмы отклика на перегрузку в пути доставки пакетов ACK являются предметом современных и будущих исследований (одним из возможных вариантов может служить снижение отправителем размера окна насыщения при получении чистого пакета ACK с кодом CE). Для современных реализаций TCP отбрасывание одного пакета ACK в общем случае оказывает пренебрежимо малое влияние на скорость передачи TCP.

7. Изменения, требуемые для IP и TCP

Требуется спецификация двух битов заголовка IP - флага ECT[1]) и флага CE[2]. Нулевое значение бита ECT говорит о том, что транспортный протокол будет игнорировать флаг CE. Это значение бита ECT используется по умолчанию. Если флаг ECT установлен (1), это говорит о том, что транспортный протокол способен принимать участие в ECN.

По умолчанию флаг CE имеет значение 0. Маршрутизатор устанавливает CE = 1 для индикации перегрузки конечным узлам. Маршрутизаторам ни в коем случае не следует сбрасывать бит CE в заголовках пакетов из 1 в 0.

В протокол TCP требуется внести три изменения - фаза согласования на этапе организации соединения для определения поддержки ECN обоими узлами и два новых флага в заголовке TCP из резервного пространства в поле флагов TCP. Флаг ECN-Echo используется получателем данных для информирования отправителя полученного пакета CE. Флаг CWR используется отправителем данных для информирования получателя о снижении размера окна насыщения.

8. Отсутствие связи с индикаторами ATM EFCI и Frame Relay FECN

Поскольку механизмы индикации насыщения в ATM и Frame Relay без связи со средним размером очереди, как основой для определения перегрузки промежуточного узла, мы полагаем, что такая индикация будет создавать избыточный шум. Реакция отправителя TCP в соответствии с данной спецификацией для ECN **не является** подходящим вариантом для такого шумного сигнала о насыщении. Мы надеемся, что механизмы ATM EFCI и Frame Relay FECN будут поэтапно развертываться в сетях ATM. Однако, если маршрутизаторы, имеющие интерфейсы в сети ATM, получат способ определения среднего размера очереди для интерфейса и станут использовать этот размер для надежного детектирования перегрузки в подсети ATM, они могут использовать уведомления ECN, описанные в данной спецификации.

Мы подчеркиваем, что транспортный уровень реагирует в плане контроля насыщения на «один» пакет с флагом CE в заголовке IP так же, как он реагировал бы на отбрасывание пакета. Таким образом, бит CE не обеспечивает достаточно хорошего соответствия сигналам, основанным на мгновенном размере очереди. Однако эксперименты с методами управления на уровне 2 (например, в коммутаторах ATM и Frame Relay) следует поощрять. Например, используя схемы типа RED (когда пакеты маркируются на основе превышения средним размером очереди заданного порога), устройства канального уровня могут обеспечивать достаточно надежную индикацию насыщения. Когда все устройства уровня 2 на пути установят принятый на этом уровне маркер возможного насыщения (например, бит EFCI для ATM или бит FECN для Frame Relay) с использованием надежного детектирования перегрузки, интерфейс маршрутизатора в сеть уровня 2 сможет транслировать такие маркеры в маркеры CE заголовков IP. Мы признаем, что в сегодняшней практике и стандартах такого не наблюдается. Однако продолжение экспериментов в этом направлении может дать информацию, которая позволит найти способ перехода от существующих механизмов канального уровня к более надежной индикации насыщения с использованием для индикации перегрузки одного бита.

9. Неподатливость конечных узлов

В этом разделе рассматривается опасность ECN для неподатливых[3] конечных узлов (т. е., узлов, которые устанавливают код ECT в передаваемых пакетах, но не реагируют на получение пакетов с кодом CE). Мы понимаем, что добавление ECN в архитектуру IP не повышает сколь-нибудь существенно общий уровень уязвимости архитектуры со стороны невосприимчивых потоков.

---

[1]ECN-Capable Transport - транспорт, поддерживающий ECN.
[2]Congestion Experienced - наблюдается насыщение.
[3]Non-compliant.

Даже для сред, не поддерживающих ECN, следует серьезно рассматривать возможность нарушений, которые могут быть вызваны неподатливыми или невосприимчивыми потоками (т. е., потоками, которые не отвечают на индикацию насыщения снижением скорости доставки через загруженный канал). Например, конечная точка может «отключить контроль насыщения», не снижая размер окна насыщения в ответ на отбрасывание пакетов. Эта проблема важна для современного состояния Internet. Ясно, что в маршрутизаторах нужно реализовать механизмы детектирования и дифференцированной трактовки пакетов из неподатливых потоков. Предполагается также, что такие методы, как сквозное планирование на уровне потока и изоляция потоков друг от друга, дифференцированные услуги или сквозное резервирование могут устранить некоторые из наиболее разрушительных эффектов от невосприимчивых потоков.

Можно сказать, что само по себе отбрасывание пакетов является средством сдерживания неподатливости, а использование ECN блокирует эту возможность. Мы утверждаем в ответ на это, что (1) поддерживающие ECN маршрутизаторы сохраняют возможность отбрасывания пакетов при сильной перегрузке и (2) даже в случаях сильной перегрузки отбрасывание пакетов не является сдерживающим неподатливость фактором.

Во-первых, поддерживающие ECN маршрутизаторы будут только маркировать пакеты (вместо их отбрасывания), пока частота маркирования достаточно мала. В периоды, когда средний размер очереди превышает верхний порог и, следовательно, потенциальная скорость маркировки пакетов будет велика, мы рекомендуем маршрутизатору отбрасывать пакеты вместо установки кода CE в заголовках.

В периоды с низкой и средней скоростью маркировки пакетов, когда поддержка ECN реализована, будет возникать некий негативный эффект для невосприимчивых потоков в виде отбрасывания пакетов вместо их маркировки. Например, для нечувствительных к задержкам потоков, использующих гарантированную доставку, при отбрасывании пакетов может наблюдаться увеличение скорости вместо ее снижения. Аналогично для чувствительных к задержкам потоков без гарантированной доставки может возрастать использование FEC в ответ на рост частоты отбрасывания пакетов, приводящее скорее к росту, чем к снижению скорости передачи. По тем же причинам мы не верим, что отбрасывание пакетов, само по себе, является эффективным средством сдерживания для неподатливости даже в средах с высокой частотой отбрасывания пакетов, когда вероятность отбрасывания делится между всеми потоками.

Было предложено несколько методов идентификации и ограничения неподатливых и невосприимчивых потоков. Добавление ECN в сетевую среду никак не усложнит разработку и развертывание таких механизмов. Во всяком случае, добавление ECN в архитектуру будет существенно упрощать работу по идентификации невосприимчивых потоков. Например, в среде с поддержкой ECN маршрутизаторы не ограничиваются информацией о том, что пакет был отброшен или получил код на данном маршрутизаторе - в таких средах маршрутизаторы могут также отмечать прибытие пакетов с кодом CE, показывающих перегрузку, встреченную пакетом на своем пути раньше.

10. Неподатливость в сети

Снижение эффективности контроля насыщения может быть обусловлено неподатливостью не только оконечных узлов, но и утратой индикации насыщения в сети. Причиной такой утраты могут быть враждебные или некорректно работающие маршрутизаторы, которые устанавливают флаг ECT в пакетах не поддерживающего ECN транспорта или «удаляют» флаг CE из прибывающих пакетов. Например, маршрутизатор, «удаляющий» флаг CE из прибывающих пакетов CE будет блокировать индикацию насыщения от нисходящих получателей. Это может приводить к отказу контроля насыщения для данного потока и, в конечном счете, отбрасыванию последующих пакетов данного потока при увеличении среднего размера очереди на перегруженном шлюзе.

Действия враждебного или работающего некорректно маршрутизатора могут также приводить к ложной индикации насыщения конечным узлам. Такие действия могут включать отбрасывание пакетов маршрутизатором или установку флага CE при отсутствии насыщения. С точки зрения контроля насыщения установка флага CE при отсутствии перегрузки не будет отличаться от необоснованного отбрасывания пакетов маршрутизатором. Сброс флага ECT для пакетов, которые будут позднее отброшены в сети может приводить к неоправданному отбрасыванию пакетов.

Проблемы, связанные с потерей индикации насыщения в инкапсулированных, отброшенных и поврежденных пакетах, рассматриваются ниже.

10.1. Инкапсулированные пакеты

При обработке битов CE и ECT в процессах инкапсуляции и декапсуляции для туннелей следует соблюдать осторожность.

При инкапсяляции пакетов нужно выполнять приведенные здесь правила для бита ECT. Во-первых, если флаг ECT в инкапсулированном (внутреннем) заголовке сброшен (0), для флага ECT в инкапсулирующем (внешнем) заголовке **должно** быть установлено значение 0. Если ECT во внутреннем заголовке имеет значение 1, для флага ECT во внешнем заголовке **следует** установить значение 1.

При декапсуляции пакета используются приведенные здесь правила для флага CE. Если бит ECT имеет значение 1 как во внешнем, так и во внутреннем заголовке, нужно использовать операцию «логическое ИЛИ» (OR) для полей CE внешнего и внутреннего заголовка (т. е., при установленном во внешнем заголовке флаге CE этот флаг должен копироваться во внутренний заголовок). Если флаг ECT в одном из заголовков сброшен (0), значение бита CE во внешнем заголовке игнорируется. Это требование в настоящее время не применяется при декапсуляции в туннелях IPsec.

Специфическим примером использования ECN с инкапсуляцией является случай использования потоком поддержки ECN для предотвращения нежелательного отбрасывания пакетов в результате перегрузки на промежуточном узле в туннеле. Такая функциональность может быть обеспечена путем копирования поля ECN из внутреннего заголовка IP во внешний при инкапсуляции и использования поля ECN во внешнем заголовке IP для установки поля ECN внутреннего заголовка IP при декапсуляции. Это позволяет маршрутизаторам по пути туннеля устанавливать бит CE в поле ECN неинкапсулированного заголовка IP для поддерживающего ECN пакета, когда маршрутизатор испытывает перегрузку.

10.2. Туннели IPsec

Протокол IPsec, как определено в [ESP, AH], не включает поле ECN заголовка IP в криптографические преобразования (в туннельном режиме поле ECN не включается во внутренний заголовок IP). Следовательно, изменение поля ECN узлом сети не оказывает влияния на сквозную защиту IPsec, поскольку изменение этого поля не может быть

зафиксировано средствами контроля целостности IPsec. В результате этого IPsec не обеспечивает какой-либо защиты от враждебного изменения поля ECN (т. е., MITM-атак[1]), поскольку враждебные изменения также не будут оказывать влияния на сквозную защиту IPsec. В некоторых средах способность изменить поле ECN без влияния на проверку целостности IPsec позволяет создавать скрытые каналы - если требуется предотвратить создание такого канала или снизить доступную для него полосу, значение поля ECN во внешнем заголовке можно обнулять на входном и выходном узлах.

Протокол IPsec в настоящее время требует, чтобы поле ECN внутреннего заголовка не менялось в процессе декапсуляции IPsec на выходе из туннеля. Это позволяет предотвратить возможность организации атак за счет изменения поля ECN в конечных точках туннеля, поскольку внесенные изменения будут отбрасываться в конце туннеля. Данный документ не меняет это требование IPsec. С учетом текущей спецификации протокола IPsec мы полагаем, что эксперименты с полем ECN в настоящее время не могут проводиться для потоков, использующих туннели IPsec.

Если спецификация IPsec в будущем позволит выходным узлам туннелей менять поле ECN внутреннего заголовка IP на основе значения поля ECN во внешнем заголовке (например, путем полного или частичного копирования поля ECN из внешнего заголовка во внутренний) или обнулять поле ECN внешнего заголовка IP в процессе инкапсуляции, эксперименты с полем ECN могут быть выполнены и для туннелей IPsec.

Обсуждение взаимодействия ECN с туннелями IPsec во многом основано на обсуждениях и документах рабочей группы Differentiated Services.

10.3. Отброшенные и поврежденные пакеты

Возникает дополнительный вопрос применительно к пакетам, в которых один маршрутизатор устанавливает флаг CE, а последующий маршрутизатор отбрасывает пакет. Для предлагаемого в этом документе использования ECN (т. е., для транспортных протоколов типа TCP, в которых отбрасывание данных является индикацией насыщения) конечные узлы детектируют отбрасывание пакетов данных и отклик (о насыщении) от обнаруживших такое отбрасывание конечных узлов имеет по крайней мере такую же силу, как отклик на получение пакета CE.

Однако транспортные протоколы типа TCP не обязательно детектируют отбрасывание любых пакетов (в частности, пакетов, содержащих лишь подтверждение ACK); например, TCP не снижает скорость доставки последующих пакетов ACK в ответ на ранее отброшенные пакеты ACK. Любые предложения по расширению ECN-Capability на такие пакеты будут приводить к возникновению проблем, таких, как маркировка пакета ACK кодом CE и последующее отбрасывание такого пакета в сети.

Аналогично, если пакет с маркировкой CE отбрасывается в сети по причине повреждения (битовые ошибки), конечным узлам следует по-прежнему вводить контроль насыщения так же, как TCP реагирует в настоящее время на отбрасывание пакета данных. Вопрос повреждения пакетов CE будет рассматриваться в любых предложениях по способам определения был пакет отброшен в результате повреждения или по причине перегрузки/переполнения буфера.

11. Обзор связанных работ

В работе [Floyd94] рассматриваются преимущества и недостатки, вносимые добавлением ECN в архитектуру TCP/IP. Как показано в основанном на моделировании сравнении, преимущество ECN заключается в предотвращении неоправданного отбрасывания пакетов для краткосрочных и чувствительных к задержкам соединений TCP. Другим преимуществом ECN является предотвращение ненужных тайм-аутов повтора передачи в TCP. В этом документе подробно рассматривается интеграция ECN с механизмами контроля насыщения TCP. Возможными недостатками ECN, отмеченными в работе, является то, что не поддерживающие ECN соединения TCP могут анонсировать себя, как ECN-совместимые, а также то, что пакеты TCP ACK, передающие сообщение ECN-Echo, могут отбрасываться в сети. Решение первого из этих вопросов был рассмотрено в разделе 8 настоящего документа, а второй решается с помощью предложений параграфа 5.1.3 для флага CWR в заголовке TCP.

В работе [CKLTZ97] описана экспериментальная реализация ECN для IPv6. Эксперименты включали реализацию ECN в существующей системе RED для FreeBSD. Было проведено множество экспериментов для демонстрации контроля за средним размером очереди в маршрутизаторе, производительности ECN для одного соединения TCP через перегруженный маршрутизатор и беспристрастности для множества одновременных соединений TCP. Один из результатов этих экспериментов заключается в том, что отбрасывание пакетов при переносе больших объемов данных может снижать производительность значительно сильнее, нежели это происходит при маркировке пакетов.

Поскольку экспериментальная реализация [CKLTZ97] в какой-то мере предшествовала разработке данного документа, она не вполне соответствует требованиям, содержащимся здесь. Например, в экспериментальной реализации не использовался флаг CWR, а вместо этого получатель передавал бит ECN-Echo в пакете ACK.

Работы [K98] и [CKLTZ98], основанные на [CKLTZ97], содержат дальнейший анализ преимуществ внедрения ECN для TCP. Заключение этих работ состоит в том, что TCP с поддержкой ECN обеспечивает некоторое повышение производительности по сравнению с TCP без ECN, потоки ECN TCP не подавляют потоков без поддержки ECN и ECN TCP обеспечивает устойчивость к двухстороннему трафику, перегрузке в обоих направлениях и множеству перегруженных шлюзов. Многочисленные эксперименты с короткими web-транзакциями показывают, что время передачи в большинстве случаев слабо зависит от применения ECN, однако иногда при использовании ECN время передачи существенно сокращалось по сравнению с передачей без использования ECN. При таких коротких передачах отбрасывание первого пакета в случае без поддержки ECN может приводить к значительному замедлению процесса за счет ожидания (до 6 сек.) тайм-аута повторной передачи.

На Web-странице ECN [ECN] приведены ссылки на другие реализации ECN находящиеся в стадии разработки.

12. Заключение

С учетом нынешних усилий по реализации RED мы полагаем, что для производителей маршрутизаторов настал момент для изучения методов реализации механизмов предотвращения перегрузки, не зависящих от отбрасывания

[1]Man-in-the-middle attack — атака с перехватом данных на пути и участием человека. *Прим. перев.*

пакетов. По мере более широкого развертывания приложений и транспорта, чувствительных к задержкам и потере одиночных пакетов (трафик в реальном масштабе времени, короткие web-транзакции и т. п.) использование потери пакетов в качестве обычного механизма индикации перегрузки представляется недостаточной мерой (во всяком случае, неоптимальной).

13. Благодарности

В подготовку этого RFC внесло свой вклад множество людей. В частности, мы выражаем свою признательность Kenjiro Cho за предложенный механизм TCP для согласования ECN-Capability, Kevin Fall за предложенный флаг CWR, Steve Blake за материал по пересчету контрольной суммы заголовка IPv4, Jamal Hadi Salim за обсуждение ECN, а также Steve Bellovin, Jim Bound, Brian Carpenter, Paul Ferguson, Stephen Kent, Greg Minshall и Vern Paxson за обсуждение вопросов безопасности. Мы также благодарим исследовательскую группу Internet End-to-End Research Group for

за полезные дмскуссии.

14. Литература

[AH] Kent, S. and R. Atkinson, "IP Authentication Header", RFC 2402[1], November 1998.

[B97] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119[2], March 1997.

[CKLT98] Chen, C., Krishnan, H., Leung, S., Tang, N., and Zhang, L., "Implementing ECN for TCP/IPv6", presentation to the ECN BOF at the L.A. IETF, March 1998, URL "http://www.cs.ucla.edu/~hari/ecn-ietf.ps".

[DIFFSERV] Nichols, K., Blake, S., Baker, F. and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474[3], December 1998.

[ECN] "The ECN Web Page", URL "http://www-nrg.ee.lbl.gov/floyd/ecn.html".

[ESP] Kent, S. and R. Atkinson, "IP Encapsulating Security Payload", RFC 2406[4], November 1998.

[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. URL "ftp://ftp.ee.lbl.gov/papers/early.pdf".

[Floyd94] Floyd, S., "TCP and Explicit Congestion Notification", ACM Computer Communication Review, V. 24 N. 5, October 1994, p. 10-23. URL "ftp://ftp.ee.lbl.gov/papers/tcp_ecn.4.ps.Z".

[Floyd97] Floyd, S., and Fall, K., "Router Mechanisms to Support End-to-End Congestion Control", Technical report, February 1997. URL "http://www-nrg.ee.lbl.gov/floyd/end2end-paper.html".

[Floyd98] Floyd, S., "The ECN Validation Test in the NS Simulator", URL "http://www-mash.cs.berkeley.edu/ns/", test tcl/test/test-all-ecn.

[K98] Krishnan, H., "Analyzing Explicit Congestion Notification (ECN) benefits for TCP", Master's thesis, UCLA, 1998, URL "http://www.cs.ucla.edu/~hari/software/ecn/ecn_report.ps.gz".

[FRED] Lin, D., and Morris, R., "Dynamics of Random Early Detection", SIGCOMM '97, September 1997. URL "http://www.inria.fr/rodeo/sigcomm97/program.html#ab078".

[Jacobson88] V. Jacobson, "Congestion Avoidance and Control", Proc. ACM SIGCOMM '88, pp. 314-329. URL "ftp://ftp.ee.lbl.gov/papers/congavoid.ps.Z".

[Jacobson90] V. Jacobson, "Modified TCP Congestion Avoidance Algorithm", Message to end2end-interest mailing list, April 1990. URL "ftp://ftp.ee.lbl.gov/email/vanj.90apr30.txt".

[MJV96] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven Layered Multicast", SIGCOMM '96, August 1996, pp. 117-130.

[RFC791] Postel, J., "Internet Protocol", STD 5, RFC 791[2], September 1981.

[RFC793] Postel, J., "Transmission Control Protocol", STD 7, RFC 793[2], September 1981.

[RFC1141] Mallory, T. and A. Kullberg, "Incremental Updating of the Internet Checksum", RFC 1141[5], January 1990.

[RFC1349] Almquist, P., "Type of Service in the Internet Protocol Suite", RFC 1349[6], July 1992.

[RFC1455] Eastlake, D., "Physical Link Security Type of Service", RFC 1455[6], May 1993.

[RFC2001] Stevens, W., "TCP Slow Start, Congestion Avoidance, Fast Retransmit, and Fast Recovery Algorithms", RFC 2001[2], January 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.

[RJ90] K. K. Ramakrishnan and Raj Jain, "A Binary Feedback Scheme for Congestion Avoidance in Computer Networks", ACM Transactions on Computer Systems, Vol.8, No.2, pp. 158-181, May 1990.

[1]Этот документ признан устаревшим и заменен RFC 4302 и RFC 4305, переводы которых имеются на сайте www.protocols.ru. *Прим. перев.*

[2]Перевод этого документа имеется на сайте www.protocols.ru. *Прим. перев.*

[3]Этот документ частично обновлен в RFC 3168 и RFC 3260. Переводы этих документов имеются на сайте www.protocols.ru. *Прим. перев.*

[4]Этот документ признан устаревшим и заменен RFC 4303 и RFC 4305, переводы которых имеются на сайте www.protocols.ru. *Прим. перев.*

[5]Этот документ обновлен в RFC 1624. *Прим. перев.*

[6]Этот документ признан устаревшим и заменен RFC 2474, перевод которого имеется на сайте www.protocols.ru. *Прим. перев.*

15. Вопросы безопасности

Вопросы безопасности рассмотрены в разделе 9.

16. Пересчет контрольной суммы заголовка IPv4

При пересчете контрольной суммы заголовка IPv4 возникает проблема с некоторыми маршрутизаторами, использующими буферизацию на выходе, поскольку большинство (если не все) операций с заголовком выполняются на входе, а решение для ECN нужно принимать локально по состоянию выходного буфера. Этой проблемы не возникает для IPv6, поскольку этот протокол не использует контрольных сумм для заголовков. Октет IPv4 TOS является последним байтом 16-битового полуслова.

В RFC 1141 [RFC1141] обсуждается нарастающее обновление контрольной суммы IPv4 после уменьшения значения поля TTL. Нарастающее обновление контрольной суммы IPv4 после установки кода CE описано ниже. Обозначим HC исходную контрольную сумму заголовка для пакета ECT(0), а HC' будет новой контрольной суммой после установки бита CE (т. е., поле ECN изменит значение с 10 на 11). Тогда контрольная сумма заголовка вычисляется путем вычитания дополнения до 1:

```
HC' = { HC - 1      HC > 1
      { 0x0000      HC = 1
```

Для расчета контрольной суммы на машинах с дополнением до двух HC' после установки флага CE будет:

```
HC' = { HC - 1      HC > 0
      { 0xFFFE      HC = 0
```

17. Обоснование для флага ECT

Необходимость введения кода ECT обусловлена тем, что развертывание ECN в сети Internet будет осуществляться поэтапно и не все транспортные протоколы и маршрутизаторы будут понимать ECN. При использовании кода ECT маршрутизатор может отбрасывать пакеты, которые не совместимы с ECN, но может «взамен» отбрасывания устанавливать код CE в пакетах, которые «поддерживают» ECN. Поскольку код ECT позволяет конечному узлу получать код CE «вместо» информации об отбрасывании пакета, это дает стимул для внедрения ECN.

Если в пакете не было кода ECT, маршрутизатор будет устанавливать код CE, как для поддерживающих, так и для не поддерживающих ECN потоков. В этом случае для конечных узлов нет стимула развертывать ECN, а также не обеспечивается путь постепенного перехода к повсеместному использованию ECN. Рассмотрим первый этап постепенного развертывания ECN, когда только часть потоков поддерживает ECN. В начале насыщения, когда скорость отбрасывания/маркировки пакетов мала, маршрутизаторы будут только устанавливать код CE, не отбрасывая пакетов. Однако понимать пакеты с кодом CE и должным образом реагировать на них будут только потоки, поддерживающие ECN. В результате поддерживающие ECN потоки будут снижать скорость, а не понимающие сигналов ECN потоки будут работать с прежним размером окна насыщения.

В этом случае возможны два варианта: (1) поддерживающий ECN поток снижает скорость, не поддерживающий ECN поток забирает освободившуюся полосу и насыщение сохраняется или (2) поддерживающий ECN поток снижает скорость, а не поддерживающий - не снижает и перегрузка возрастает, пока маршрутизатор не переходит от маркировки пакетов кодом CE к отбрасыванию пакетов. Хотя второй вариант не вполне беспристрастен, поддерживающий ECN поток в этом получает некоторые преимущества, поскольку увеличившееся насыщение заставляет маршрутизатор перейти в режим отбрасывания пакетов.

Поток, анонсирующий свою поддержку ECN, но не отвечающий на код CE, функционально эквивалентен потоку, в котором выключен контроль насыщения, как обсуждалось в параграфах 8 и 9.

Таким образом, среда, где часть потоков поддерживает ECN, но эти потоки не имеют механизма для индикации такой поддержки маршрутизаторам, будет менее эффективной и более пристрастно реагировать на перегрузки, что явится стимулом для конечных узлов к развертыванию ECN.

18. Зачем использовать два бита в заголовке IP?

Необходимость индикации ECT в заголовке IP понятна, но остается вопрос о возможности использования для кодов ECT (транспорт с поддержкой ECN) и CE (обнаружено насыщение) одного бита в заголовке. Такое однобитовое представление предложено в работе [Floyd94]. Одно значение «ECT, но без CE» будет представлять поддерживающий ECN транспорт, а другое - «CE или без ECT» будет представлять факт насыщения или транспорт без поддержки ECN.

Различие между однобитовой и двухбитовой реализацией возникает для пакетов, проходящих через множество перегруженных маршрутизаторов. Рассмотрим пакет с кодом CE, который приходит на второй перегруженный маршрутизатор и выбирается системой активного управления очередью на маршрутизаторе для маркировки или отбрасывания. В однобитовом варианте второму перегруженному маршрутизатору остается только один вариант - отбросить пакет с кодом CE, поскольку этот маршрутизатор не может отличить пакет с кодом CE от пакета без ECT. В двухбитовом варианте второй перегруженный маршрутизатор может отбросить пакет с кодом CE или переслать его дальше, сохранив код CE.

Другое различие между однобитовыми и двухбитовыми реализациями заключается в том, что в однобитовом случае получатели не могут различить пакеты CE и non-ECT в одном потоке. Таким образом, в однобитовой реализации поддерживающий ECN отправитель будет давать получателю однозначную индикацию поддержки ECN. У отправителя остается возможность показать поддержку ECN в заголовке транспортного уровня. Другим вариантом является функциональное ограничение для однобитовых реализаций, в соответствии с которым отправитель трактует **все** переданные им пакеты как поддерживающие или не поддерживающие ECN. Для транспортных протоколов с групповой адресацией такая однозначная индикация будет видная получателям, присоединившимся к действующему multicast-сеансу.

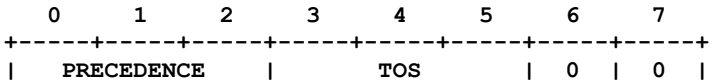
Еще одним преимуществом двухбитового варианта является повышенная отказоустойчивость. Наиболее критический момент, описанный в разделе 8, заключается в том, что по умолчанию следует указывать не поддерживающий ECN транспорт. В двухбитовом варианте для реализации этого требования достаточно просто устанавливать по умолчанию код not-ECT. В однобитовом варианте для выполнения этого требования следует устанавливать код "CE или ECT".

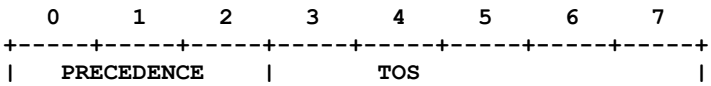
Этот вариант менее понятен и, возможно, более открыт для некорректных реализаций на конечных узлах или маршрутизаторах.

Хотя в целом однобитовая реализация вполне допустима, она имеет ряд существенных недостатков по сравнению с двухбитовым вариантом. Во-первых, функциональность однобитового варианта существенно ограничена в плане трактовки пакетов с кодом CE на втором перегруженном маршрутизаторе. Во-вторых, однобитовый вариант требует передачи дополнительной информации в заголовке транспортного уровня пакетов из поддерживающего ECN потока (функциональность двухбитового варианта просто переносится на транспортный уровень) или понимания со стороны отправителей поддерживающих ECN потоков того, что получатели должны быть способны a-priori определить какие пакеты поддерживают ECN, а какие не поддерживают. В-третьих, однобитовая реализация потенциально более открыта для ошибок со стороны некорректных реализаций, которые могут по умолчанию устанавливать неверное значение бита ECN. Мы полагаем, что перечисленные ограничения обеспечивают достаточные основания использования дополнительного бита в заголовке IP для кодов ECT.

19. Ретроспектива использования октета IPv4 TOS

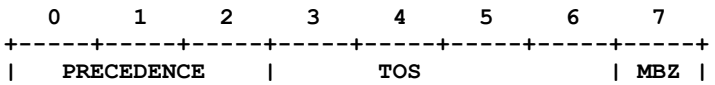
RFC 791 [RFC791] определяет октет ToS[1] в заголовке IP. В RFC 791 биты 6 и 7 октета ToS отмечены, как резервные (Reserved for Future Use), и указано, что они имеют нулевое значение. Первые два поля октета ToS определены в документе, как Precedence (предпочтения) и Type of Service (тип обслуживания - TOS).





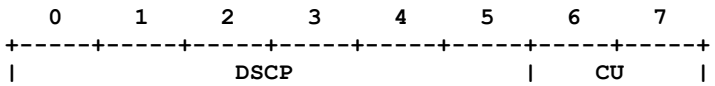
RFC 1122 включает биты 6 и 7 в поле ToS, не обсуждая конкретного их использования (см рисунок слева).

Октет ToS заголовка IPv4 был заново определен в RFC 1349 [RFC1349], как показано ниже.



Бит 6 поля TOS был определен в RFC 1349, как «Minimize Monetary Cost[2]». В дополнение к полям Precedence и TOS было определено поле MBZ[3], которое в настоящее время не используется. В RFC 1349 отмечено, что отправитель дейтаграмм устанавливает в поле MBZ нулевое значение, если не используется экспериментальных протоколов с иной трактовкой этого бита.

RFC 1455 [RFC 1455] определяет экспериментальный стандарт, использующий все четыре бита поля TOS для запроса гарантированного уровня защиты канала.



Документы RFC 1349 и RFC 1455 были отменены RFC 2474 «Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers[4]» [DIFFSERV], в котором биты 6 и 7 поля DS были указаны, как неиспользуемые (CU[5]). Первые шесть битов поля DS трактуются, как коды дифференцированного обслуживания (DSCP[6]). Формат поля показан на рисунке.

Поскольку история еще не завершена, определение поля ECN в данном документе не гарантирует совместимости со всеми предшествующими вариантами использования этих двух битов. Помехи, которые могут создавать некорректно работающие маршрутизаторы, включают «удаление» кода CE для поддерживающих ECN пакетов, которые поступили на маршрутизатор с установленным флагом CE, и установку кода CE при отсутствии перегрузок. Эти вопросы рассмотрены в параграфе «Неподатливость в сети».

Нарушения в работе поддерживающей ECN среды, которые могут создаваться не совместимыми с ECN конечными узлами, передающими пакеты с установленным кодом ECT, рассмотрены в параграфе «Неподатливость конечных узлов».

Адреса авторов

K. K. Ramakrishnan

AT&T Labs. Research

Phone: +1 (973) 360-8766

EMail: kkrama@research.att.com

URL: http://www.research.att.com/info/kkrama

Sally Floyd

Lawrence Berkeley National Laboratory

Phone: +1 (510) 486-7518

EMail: floyd@ee.lbl.gov

[1]Type of Service — тип обслуживания.
[2]Минимизация финансовых расходов.
[3]Must be zero — должно иметь нулевое значение.
[4]Определение поля дифференцированного обслуживания (DS) в заголовках IPv4 и IPv6.
[5]Currently Unused — в настоящее время не используются.
[6]Differentiated Services CodePoint.

URL: http://www-nrg.ee.lbl.gov/floyd/

Перевод на русский язык

Николай Малых

nmalykh@protocols.ru

Полное заявление авторских прав