ШПД: Need for Speed гигабит’s

"И да, я все еще мечтаю об ускорении загрузки!" - так в свое время говорил один из главных героев в первой серии культового сериала "Теория большого взрыва". Здесь ускорение загрузки по смыслу равно увеличению пропускной способности канала доступа в интернет. Восемь сезонов спустя, а это, для какого-то долгих, а для кого-то как один миг, восемь лет, - в сетях операторов, по крайней мере, российских, - внимание на развитие, как правило, все также вертится, подобно карусели, вокруг того, какая ширина канала отдана пользователям. На примере Краснодара за это время пропускная способность доступная для абонентов в жилом секторе от первых псевдо-безлимитных предложений на скорости 64К(бит/сек) напрямую шагнула к пределу 100М. Если идти также быстро дальше, недалек тот день, когда декларируемая скорость передачи на входе в квартиру пользователя подойдет к следующему рубежу, а операторы задумаются о предоставлении полосы больше гигабита.

В то, что операторы со временем дадут пользователям гигабитные, а надо будет и десяти гигабитные каналы (а надо будет очень, и более) – сомневаться, не приходится. Вот только смогут ли пользователи оценить, в какой уже раз, еще "более быстрый интернет", когда его скорость растет лишь путем увеличения доступной полосы на порту доступа? Есть много причин, почему банальное увеличение полосы с 5 до 100М не делает интернет, на взгляд привычного для пользователя опыта, в 20 раз быстрее. На самом деле более широкая полоса никогда не будет лишней, особенно когда вам надо выполнить задачи, которые требуют передачи значительных объемов данных: загрузка больших файлов или потоковое видео в высоком разрешении. Чтобы скачать файл размером 50 Гбайт при скорости подключения 2М, потребуется больше часа, и всего три минуты, если речь идет о подключении на скорости 50М.

Однако при загрузке небольших файлов, как это происходит при открытии типичной веб-страницы, состоящей из множества отдельных элементов общим объемом в пару Мбайт, разница во времени для пользователя, будь это две или одна секунда, станет уже не так заметна. На самом деле при увеличении скорости соединения с 5 до 10М, как показывают тесты Akamai на 25 наиболее популярных страницах среди пользователей США (тесты проводились в 2009 году), эта разница буквально составляет только 83 миллисекунды, соответственно уменьшение с 1443 до 1360 мс. Вряд ли даже самый из нетерпеливых пользователей за это время успеет моргнуть, если это конечно не какой-нибудь супергерой, способный двигаться на сверхчеловеческих скоростях.



Latency per Bandwith, More Bandwidth Doesn’t Matter

На самом деле более широкая полоса означает только, что за определенное время можно получить больше данных, но никак не быстрее. Для аналогии можно представить маршрут в сети как автостраду, где пакеты, то есть машины, идут в среднем с одной скоростью. Можно легко увеличить ее пропускную способность, добавив дополнительные полосы для движения. За определенный момент времени через нее пройдет больше машин. Но если требуется быстро доставить товар, отправив за ним транспорт, пропускная способность здесь мало поможет. Чтобы доехать до места, забрать груз и доставить его обратно, - важно будет время, затраченное на дорогу, которое зависит от скорости и расстояния, а если вернуться обратно в мир телекома, определяется временем приема-передачи, иначе задержкой или RTT (Round-Trip Time). Когда пользователь "кликает" в окне своего браузера, прежде чем начнется загрузка элементов страницы, сначала отправляется пакет с запросом GET для каждого из них, а за мгновение до этого потребуется установить соединение по протоколу TCP.

В 80-ые года прошлого века создатели интернета были сосредоточены на надежности передачи данных, а скорость являлась вопросом второстепенным. Сеть тогда создавалась для гарантии доставки сообщений в обстановке тотального разрушения, а не для просмотра фильмов и ТВ-шоу, игр, загрузки музыки или обновления программного обеспечения. Для надежной доставки данных предназначен протокол TCP (Transmission Control Protocol), который является основным протоколом для передачи данных в интернете. В том числе для протокола на уровень выше, HTTP (Hyper Text Transfer Protocol), поверх которого в свою очередь работают великое множество приложений, определяемых как веб или Всемирная паутина.

Надежность TCP заключается в получении подтверждения перед отправкой данных, что ранее переданный пакет или сегмент сразу из нескольких пакетов (окно TCP) успешно получен. Увеличивая размер окна, то есть объем передаваемых данных, не дожидаясь подтверждения их получения, - увеличивается скорость передачи. Это как перевезти из одного места в другое двести мешков цемента на Газели или в багажнике своего легкового автомобиля. Однако когда возникают потери пакетов, в частности из-за перегрузки на каком либо из участков сети, перед повторной передачей набора данных, где подтверждения об их доставке не получено, размер окна уменьшается почти в два раза и соответствующе снижается скорость передачи. По аналогии, попробуйте приготовить пищу на плите, которая периодически нагревается на полную мощность, а затем остывает. Это будет еще тем наслаждением...

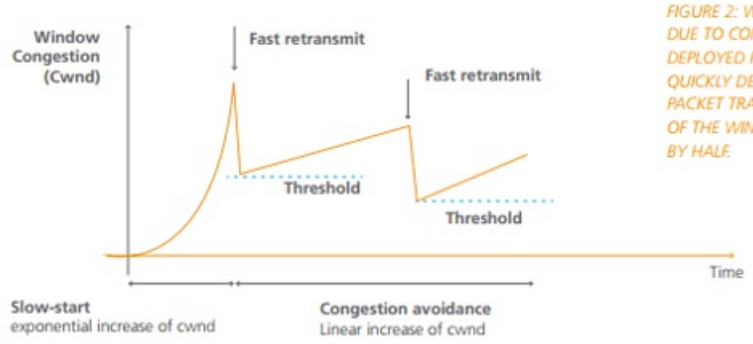
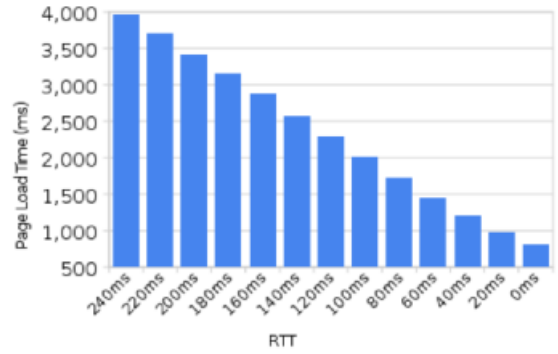


FIGURE 2: WHEN THERE IS PACKET LOSS DUE TO CONGESTION, THE ALGORITHM DEPLOYED FOR CONGESTION CONTROL QUICKLY DECREASES THE RATE OF PACKET TRANSFER BY CUTTING THE SIZE OF THE WINDOW CONGESTION ALMOST BY HALF.

[Improving Online Video Quality and Accelerating Downloads]

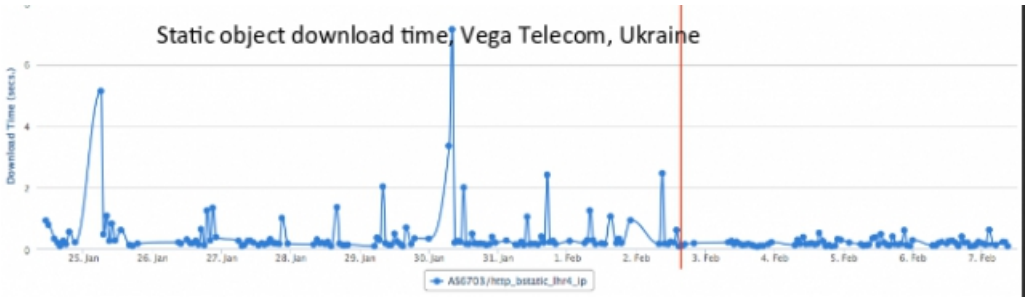
В общем, в оценке производительности сети уравнение состоит как минимум из обеих слагаемых: пропускной способности и задержки. Последняя переменная здесь, хотя по сложившейся на сегодня практике пользователи в массе своей мало знают о второй составляющей "быстрого" интернета, за исключением вероятно геймеров – играет не менее, а зачастую и более важную роль для итогового результата. Это хорошо показывает график соотношения времени загрузки типичной веб-страницы по отношению к задержке. При равной пропускной способности на каждые 20 мс уменьшения RTT, время загрузки страницы улучшается линейно. Причины для этого есть: в среднем веб-страница состоит из множества небольших элементов, для загрузки которые надо открыть много соединений, а производительность протокола TCP, в свою очередь, сильно зависит от времени прохождения пакета на маршруте от отправителя к получателю и обратно.



Page Load Time As RTT Decreases

Как видно, удвоение пропускной способности без уменьшения RTT оказывает минимальное воздействие на взгляд привычного для пользователя опыта. В тоже время уменьшение времени задержки приема-передачи, независимо от текущей пропускной способности, будет она выше, или даже немного ниже, - делает путешествия по вебу значительно удобнее. Поэтому в давно начатой операторами игре, как дать свои абонентам более быстрый интернет, надо вероятно перестать "бороться" между собой за пользователя, обещая ему только "горы" Мбит. Пора уже возможно задаться вопросом, как продавать ему, пользователю, миллисекунды задержки. Благо "отстроится" в маркетинге от соперников здесь будет проще. Всегда можно найти уникальные сервисы, где именно на своей сети время прохождения пакетов туда и обратно, по сравнению с другими, кто из конкурентов вас окружает, будет меньше.

Отсюда остается только шаг до пиринга, прямому обмену трафиком между двумя или сразу несколькими сетями, а также точкам обмена (коих в России, кстати, тоже хватает), как удобному инструменту для реализации пиринговых взаимоотношений. Впрочем, зачем покупать пиринг на аиксах будет уже немного другой историей...



Avg. Download Time Distribution, Is peering really faster? Let the data speak for its