



МЕЖДУНАРОДНЫЙ СОЮЗ ЭЛЕКТРОСВЯЗИ

МСЭ-Т

СЕКТОР СТАНДАРТИЗАЦИИ
ЭЛЕКТРОСВЯЗИ МСЭ

G.1030

(11/2005)

СЕРИЯ G: СИСТЕМЫ И СРЕДА ПЕРЕДАЧИ,
ЦИФРОВЫЕ СИСТЕМЫ И СЕТИ

Качество обслуживания и технические
характеристики – Общие и связанные с пользователем
аспекты

**Оценка сквозного качества работы в IP-сетях
для приложений передачи данных**

Рекомендация МСЭ-Т G.1030

РЕКОМЕНДАЦИИ МСЭ-Т СЕРИИ G
СИСТЕМЫ И СРЕДА ПЕРЕДАЧИ, ЦИФРОВЫЕ СИСТЕМЫ И СЕТИ

МЕЖДУНАРОДНЫЕ ТЕЛЕФОННЫЕ СОЕДИНЕНИЯ И ЦЕПИ	G.100–G.199
ОСНОВНЫЕ ХАРАКТЕРИСТИКИ, ОБЩИЕ ДЛЯ ВСЕХ АНАЛОГОВЫХ СИСТЕМ ПЕРЕДАЧИ	G.200–G.299
ИНДИВИДУАЛЬНЫЕ ХАРАКТЕРИСТИКИ МЕЖДУНАРОДНЫХ ВЧ-СИСТЕМ ТЕЛЕФОННОЙ СВЯЗИ ПО МЕТАЛЛИЧЕСКИМ ЛИНИЯМ	G.300–G.399
ОБЩИЕ ХАРАКТЕРИСТИКИ МЕЖДУНАРОДНЫХ СИСТЕМ ТЕЛЕФОННОЙ СВЯЗИ НА ОСНОВЕ РАДИОРЕЛЕЙНЫХ ИЛИ СПУТНИКОВЫХ ЛИНИЙ И ИХ СОЕДИНЕНИЕ С МЕТАЛЛИЧЕСКИМИ ПРОВОДНЫМИ ЛИНИЯМИ	G.400–G.449
КООРДИНАЦИЯ РАДИОТЕЛЕФОНИИ И ПРОВОДНОЙ ТЕЛЕФОНИИ	G.450–G.499
ХАРАКТЕРИСТИКИ СРЕДЫ ПЕРЕДАЧИ	G.600–G.699
ЦИФРОВОЕ ОКОНЕЧНОЕ ОБОРУДОВАНИЕ	G.700–G.799
ЦИФРОВЫЕ СЕТИ	G.800–G.899
ЦИФРОВЫЕ УЧАСТКИ И СИСТЕМА ЦИФРОВЫХ ЛИНИЙ	G.900–G.999
КАЧЕСТВО ОБСЛУЖИВАНИЯ И ТЕХНИЧЕСКИЕ ХАРАКТЕРИСТИКИ – ОБЩИЕ И СВЯЗАННЫЕ С ПОЛЬЗОВАТЕЛЕМ АСПЕКТЫ	G.1000–G.1999
ХАРАКТЕРИСТИКИ СРЕДЫ ПЕРЕДАЧИ	G.6000–G.6999
ПЕРЕДАЧА ДАННЫХ ПО ТРАНСПОРТНЫМ СЕТЯМ – ОБЩИЕ ПОЛОЖЕНИЯ	G.7000–G.7999
ETHERNET И АСПЕКТЫ ТРАНСПОРТИРОВКИ СООБЩЕНИЙ	G.8000–G.8999
СЕТИ ДОСТУПА	G.9000–G.9999

Для получения более подробной информации просьба обращаться к перечню Рекомендаций МСЭ-Т.

Рекомендация МСЭ-Т G.1030

Оценка сквозного качества работы в IP-сетях для приложений передачи данных

Резюме

В настоящей Рекомендации описывается набор инструментов для получения параметров качества работы IP-сети, оценки качественных показателей пользовательских приложений и применения моделей восприятия для измерения степени удовлетворённости потребителей сквозным качеством.

Воспринимаемое пользователем качество работы приложений передачи данных в сетях пакетной передачи зависит от множества факторов, включая сквозное качество работы пакетной сети, зависимости приложения от сети связи, качество работы терминалов и других устройств, не контролируемых оператором(ами) сети связи, а также задач пользователя и степени его взаимодействия с данным приложением. Для гарантированного удовлетворения потребностей пользователей разработчики сети учитывают эти факторы. После того, как дана оценка качества работы приложения, для интерпретации уровня достигнутого сквозного качества могут применяться модели восприятия.

В настоящей Рекомендации предполагается, что читатель сможет предоставить хоть какой-нибудь уровень информации о вышеперечисленных ключевых факторах и затем использовать описанный здесь набор инструментов для оценки сквозного качества.

Источник

Рекомендация МСЭ-Т G.1030 была утверждена 29 ноября 2005 года 12-й Исследовательской комиссией МСЭ-Т (2005–2008 гг.) в соответствии с процедурой, изложенной в Рекомендации МСЭ-Т А.8.

ПРЕДИСЛОВИЕ

Международный союз электросвязи (МСЭ) является специализированным учреждением Организации Объединенных Наций в области электросвязи. Сектор стандартизации электросвязи МСЭ (МСЭ-Т) – постоянный орган МСЭ. МСЭ-Т отвечает за изучение технических, эксплуатационных и тарифных вопросов и за выпуск Рекомендаций по ним с целью стандартизации электросвязи на всемирной основе.

На Всемирной ассамблее по стандартизации электросвязи (ВАСЭ), которая проводится каждые четыре года, определяются темы для изучения Исследовательскими комиссиями МСЭ-Т, которые, в свою очередь, вырабатывают Рекомендации по этим темам.

Утверждение Рекомендаций МСЭ-Т осуществляется в соответствии с процедурой, изложенной в Резолюции 1 ВАСЭ.

В некоторых областях информационных технологий, которые входят в компетенцию МСЭ-Т, необходимые стандарты разрабатываются на основе сотрудничества с ИСО и МЭК.

ПРИМЕЧАНИЕ

В настоящей Рекомендации термин "администрация" используется для краткости и обозначает как администрацию электросвязи, так и признанную эксплуатационную организацию.

Соблюдение положений данной Рекомендации носит добровольный характер. Однако в Рекомендации могут содержаться определенные обязательные положения (например, для обеспечения возможности взаимодействия или применимости), и соблюдение положений данной Рекомендации достигается в случае выполнения всех этих обязательных положений. Для выражения необходимости выполнения требований используется синтаксис долженствования и соответствующие слова (такие, как "должен" и т. п.), а также их отрицательные эквиваленты. Использование этих слов не предполагает, что соблюдение положений данной Рекомендации является обязательным для какой-либо из сторон.

ПРАВА ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ

МСЭ обращает внимание на вероятность того, что практическое применение или реализация этой Рекомендации может включать использование заявленного права интеллектуальной собственности. МСЭ не занимает какую бы то ни было позицию относительно подтверждения, обоснованности или применимости заявленных прав интеллектуальной собственности, независимо от того, отстаиваются ли они членами МСЭ или другими сторонами вне процесса подготовки Рекомендации.

На момент утверждения настоящей Рекомендации МСЭ не получил извещение об интеллектуальной собственности, защищенной патентами, которые могут потребоваться для выполнения этой Рекомендации. Однако те, кто будет применять Рекомендацию, должны иметь в виду, что это может не отражать самую последнюю информацию, и поэтому им настоятельно рекомендуется обращаться к патентной базе данных БСЭ.

© ITU 2006

Все права сохранены. Никакая часть данной публикации не может быть воспроизведена с помощью каких-либо средств без письменного разрешения МСЭ.

СОДЕРЖАНИЕ

	Стр.	
1	Общие положения и сфера применения	1
1.1	Оценка качества работы сети	1
1.2	Оценка качества работы приложения.....	2
1.3	Модели восприятия	2
1.4	Применение моделей в процессе оценки сквозного качества работы.....	2
2	Справочная литература.....	3
3	Сокращения	3
4	Оценка QoS конечного пользователя.....	3
Приложение А – Субъективная модель для приложений поиска и просмотра информации в веб		4
A.1	Общие положения.....	4
A.2	Введение	4
A.3	Субъективные эксперименты по оценке качества поиска и просмотра информации в веб и их результаты.....	5
A.4	Результаты моделирования.....	8
A.5	Воспринимаемое качество сеансов связи с просмотром одной веб-страницы и единичных событий (не стандарт)	11
A.6	Справочная литература к Приложению А.....	12
Дополнение I – Оценки качества работы сети в условиях ограниченной информации.....		13
I.1	Введение	13
I.2	Эталонное соединение	13
I.3	Комбинация значений оценок качества пакетной передачи	13
I.4	Полоса пропускания "узкого места"	14
I.5	Время соединения.....	17
I.6	Пример приложения: транзакция НТТР	18
I.7	Резюме	20
I.8	Справочная литература к Дополнению I.....	20

Введение

Воспринимаемое пользователем качество работы приложений передачи данных в пакетных сетях зависит от множества факторов, некоторые ключевые факторы перечислены ниже:

- 1) Сквозное качество работы пакетной сети (например, удобство соединения, потери и задержка пакетов, пропускная способность передачи пакетов), с учетом архитектуры сети и объема трафика пользователя. Качественные показатели сети часто являются главной составляющей сквозного качества передачи.
- 2) Зависимость приложения от сети связи (например, количество приемов-передач пакетов, необходимых для завершения сеанса связи, управление потоками для передачи данных) и других элементов сети, предоставляющих сервисы для данного приложения (например, сервера доменных имен).
- 3) Качество работы вспомогательных устройств, не контролируемых оператором(ами) сети (например, оборудование пользователя, хосты).
- 4) Задачи пользователя и степень его взаимодействия с данным приложением.

Для гарантированного удовлетворения потребностей пользователей разработчики сети учитывают эти факторы. Модель работы приложения передачи данных должна включать в себя максимально возможное число таких факторов. После того, как дана оценка качества работы приложения, для интерпретации уровня достигнутого сквозного качества могут применяться модели восприятия.

В настоящей Рекомендации предполагается, что читатель сможет предоставить хоть какой-нибудь уровень информации о вышеперечисленных ключевых факторах и затем использовать набор инструментов (многие из которых описаны в данной Рекомендации) для оценки сквозного качества работы.

Рекомендация МСЭ-Т G.1030

Оценка сквозного качества работы в IP-сетях для приложений передачи данных

1 Общие положения и сфера применения

В данной Рекомендации описывается процесс оценки сквозного качества работы приложений, работающих в IP-сетях, с использованием:

- Качественных показателей работы рассматриваемой IP-сети, полученных на основе измерений или по результатам моделирования сети.
- Спецификаций рассматриваемого приложения, в части протокола(ов) управления с определенными возможностями, или модели приложения, в которой в качестве исходных данных используются качественные показатели работы сети и качественные показатели работы абонентского оборудования, и которая создает в результате ключевые качественные параметры работы приложения (например, время загрузки файла).
- Модели восприятия, рассматриваемых приложений для интерпретации качественных показателей работы приложения и оценки качества, которое будет обеспечено для широкого пользования.

На рисунке 1 показан общий процесс получения оценки сквозного качества работы.

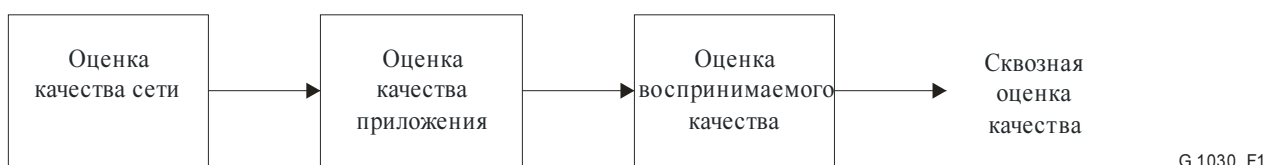


Рисунок 1/G.1030 – Процесс получения оценки сквозного качества работы

Отметим, что в некоторых случаях, например, когда моделирование позволяет измерить качественные показатели конкретного сеанса связи или некоторого количества таких сеансов, шаги для получения оценки качества работы сети и приложения могут быть объединены.

1.1 Оценка качества работы сети

Качество работы сети можно оценить в значениях параметров качества передачи пакетов, определенных в Рекомендации МСЭ-Т Y.1540, и иных соответствующих стандартах (например, RFC, разработанных Рабочей группой IETF по измерению качества IP). Кроме обычных измерений качества односторонней передачи, наиболее полную оценку качества работы приложения иногда можно получить с применением показателей двусторонней передачи.

Существует два основных источника информации о качестве работы сети – измерение и моделирование.

Измерения в сети позволяют рассматривать сеть как "черный ящик", и получать информацию, которая может быть полезна на остальных этапах цепи моделирования. Однако, существует несколько важных моментов для проведения измерений, включая:

- 1) В некоторых случаях порядок передачи должен соответствовать рассматриваемому приложению. Например, в протоколе TCP управление потоками учитывает состояние сети и стремится в узком месте линии передачи выстроить очередь пакетов, делая задержку больше, чем она была бы измерена в других условиях.
- 2) Нагрузка для реальной работы (не для измерений) по размеру и типу должна быть аналогична условиям, при которых должны применяться оценки качественных показателей. Измерения, выполненные в сети без нагрузки, не имеют особой пользы.

Моделирование сети может позволить получить необходимые характеристики качественных показателей, когда сеть еще не полностью построена (например, узлы и линии уже готовы, но наиболее сложный протокол еще не работает), или когда невозможно получить ключевые параметры

для измерений. Существует огромный выбор инструментов моделирования, включая коммерческие продукты и продукты, предназначенные для общего пользования. Для эффективного применения инструментов моделирования необходима высокая квалификация и обширная информация о рассматриваемой сети. В качестве альтернативы такому скрупулезному моделированию сети, в Дополнении I описаны упрощенные методы, которые можно использовать для оценки сквозного качества работы IP-сети. Однако точность этого метода будет существенно зависеть от точности имеющейся информации.

1.2 Оценка качества работы приложения

Модели приложения берут в качестве исходных данных оценки качества работы сети и информацию, описывающую качество работы прикладных устройств, и получают на выходе одно или несколько ключевых параметров качества работы приложения.

Одна такая модель качества работы приложения для телефонных пакетных приложений, типичных для операций с кредитными картами и других действий в торговых точках описана в Рекомендации МСЭ-Т G.1040.

Для продолжительных сеансов передачи файлов с использованием сервиса надежной передачи битов протокола TCP модели, описанные в Дополнении I, дают точные результаты, при условии, что исходные данные не содержат ошибок.

1.3 Модели восприятия

В настоящей Рекомендации представлены модели восприятия, предназначенные для получения желаемых оценок сквозного качества работы.

В Приложении А описана модель для поиска и просмотра информации в веб.

Когда появятся другие модели, они будут введены в данную Рекомендацию в виде Приложений.

1.4 Применение моделей в процессе оценки сквозного качества работы

На рисунке 2 показаны различные варианты процесса оценки сквозного качества работы приложений в IP-сетях. На рисунке видно, что существует много возможностей для выполнения оценки, хотя, на практике при выполнении оценки необходимо комбинировать возможности, которые соответствуют цели сквозной оценки (или какой-либо еще).



ПРИМЕЧАНИЕ. – В Рекомендации МСЭ-Т G.1040 описывается вклад, вносимый сетью в качество работы приложений транзакции, и не дается полная оценка качества работы приложения, хотя в ней дается взгляд на качество работы сети с точки зрения приложения.

Рисунок 2/G.1030 – Описание получения оценки сквозного качества работы IP-сети

2 Справочная литература

В нижеследующих Рекомендациях МСЭ-Т и в других документах содержатся положения, которые, с помощью ссылки в настоящем тексте, составляют положения настоящей Рекомендации. На время публикации указанные здесь издания были действительными. Все Рекомендации и другие документы постоянно пересматриваются; поэтому всем пользователям данной Рекомендации настоятельно рекомендуется по возможности использовать последние издания перечисленных ниже Рекомендаций и других документов. Список действующих в настоящий момент Рекомендаций МСЭ-Т регулярно публикуется. Ссылка в настоящей Рекомендации на какой-либо документ не придает этому отдельному документу статуса рекомендации.

- ITU-T Recommendation G.1010 (2001), *End-user multimedia QoS categories*.
- Рекомендация МСЭ-Т G.1040 (2006 г.), *Доля сети в общем времени выполнения транзакции*.
- ITU-T Recommendation Y.1540 (2002), *Internet protocol data communication service – IP packet transfer and availability performance parameters*.

3 Сокращения

IP	Протокол Интернет
QoS	Качество обслуживания
TCP	Протокол управления передачей

4 Оценка QoS конечного пользователя

После получения оценки сквозных физических качественных показателей, разработчик сети обязательно должен оценить QoS конечного пользователя, учитывая аспекты восприятия.

В Рекомендации МСЭ-Т G.1010 приведены категории QoS для приложений передачи данных, а также для приложений передачи речи и видео. Кроме того, в Приложении А описывается субъективная модель для оценки воспринимаемого пользователем качества работы приложений поиска и просмотра информации в веб. Она позволяет выполнить приблизительную оценку на основании измеренных/запланированных значений сквозной задержки в терминах мнения пользователя.

Приложение А

Субъективная модель для приложений поиска и просмотра информации в веб

А.1 Общие положения

В данном Приложении описывается модель определения соответствия между временем отклика или загрузки, измеренного в сети или рассчитанного исходя из времени, затрачиваемого на HTTP транзакцию, и воспринимаемым качеством сеанса связи поиска и просмотра информации в веб. Эта модель основана на экспериментах, в которых изменялись значения времени отклика или загрузки в ходе сеанса связи поиска и просмотра информации в веб [1], [2]. Область применения этой модели в настоящее время ограничена сеансами связи поиска и просмотра информации в веб, которые состоят из двух этапов. Первый – во время которого выполняется запрос на поиск, второй – во время которого отображается полученная страница. Кроме того, описывается простое расширение для единичных событий, в котором моделируется ожидание одной-единственной страницы.

А.2 Введение

В ходе моделирования воспринимаемого качества работы приложений поиска и просмотра информации в веб важно учитывать тот факт, что в оценке воспринимаемого качества главным фактором будет *ожидаемое* максимальное время сеанса связи. Если ожидается, время сеанса связи составит 100 секунд, то 10-секундный сеанс связи будет восприниматься как более высококачественный, чем он был бы воспринят в случае ожидания того, что связь займет 1 секунду. Следовательно, модель реализует подход на основании продолжительности содержания и использует три различных шкалы времени 6, 15 и 60 секунд, соответствующих быстрой, средней и медленной перекачке контекста в сети, соответственно.

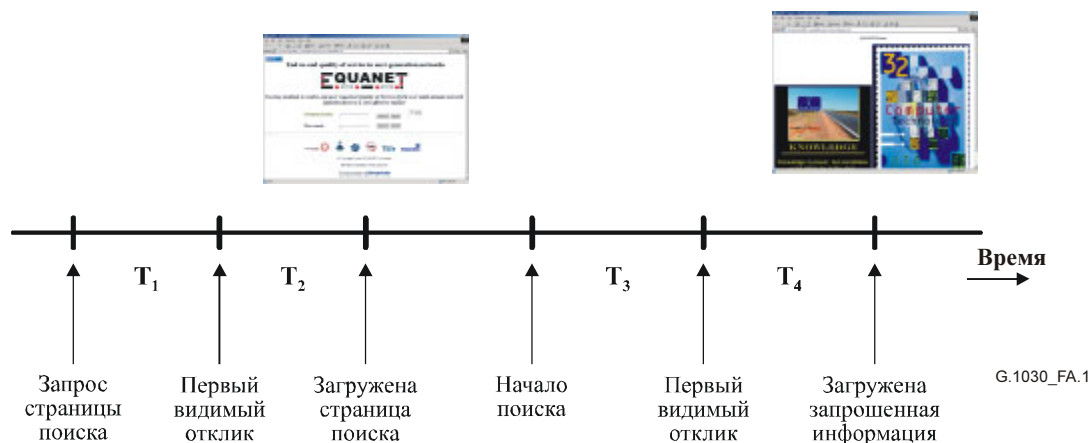
В общем случае, восприятие качества работы, связанное со временем отклика, можно классифицировать в соответствии с тремя областями восприятия [3]:

- 1) **Мгновенная реакция:** время в 0,1 секунду примерно соответствует пределу ощущения пользователя, что система реагирует мгновенно, это очень важный предел для разговорных услуг (например, чат).
- 2) **Непрерывная работа:** время в 1,0 секунду примерно соответствует пределу ощущения пользователем непрерывности потока, даже, несмотря на то, что у пользователя возникает ощущение работы через посредника, это очень важный предел для интерактивных услуг (например, сетевых игр).
- 3) **Потери внимания:** время в 10 секунд примерно соответствует пределу времени, в течение которого пользователь фокусирует внимание на диалоге. В течение более длительных задержек, ожидая пока компьютер закончит передачу, пользователи начинают выполнять другие задачи, то есть им потребуется извещение о том, когда ожидается завершение выполнения компьютером данной задачи. Извещение во время задержки особенно важно, если имеется вероятность, время отклика будет различным, поскольку пользователи при этом не знают, чего ожидать.

Что касается времени загрузки файлов, пользователи стремятся адаптировать свои суждения о качестве работы в соответствии с ожидаемым временем загрузки [4]. Когда пользователи информированы об ожидаемом времени загрузки, они готовы согласиться с тем, что время загрузки будет достаточно большим.

Модель, приведенная в настоящей Рекомендации, описывает соотношения между различными значениями времени отклика и загрузки в течение сеансов связи поиска и просмотра информации в веб и соответствующим воспринимаемым качеством поиска и просмотра информации в веб для данного *максимального* времени сеансов связи для данной конфигурации сети и системы. Данная модель применима к широкому диапазону конфигурация сетей и систем, а также к услугам поиска и просмотра информации в веб для самых разных пользователей. Субъективные эксперименты, которые послужили основой для данной модели, имитировали с максимально возможной точностью реальную ситуацию поиска и просмотра информации в веб. При разработке модели использовались три субъективных эксперимента поиска и просмотра информации в веб с временными рамками в 6,

15 и 60 секунд, представляющие быструю, среднюю и медленную работу сети, соответственно. В каждом сеансе связи пользователь сначала открывал страницу поиска, а затем страницу с результатами поиска. На рисунке А.1, показана ось времени такого сеанса связи. Первые два интервала времени T_1 и T_2 являются интервалами не интерактивных отклика и загрузки страницы поиска. Вторые два интервала времени T_3 и T_4 , являются интервалами интерактивных отклика и загрузки страницы с результатами поиска.



T_1 – интервал не интерактивного отклика, он регулировался с использованием языка Java. T_2 – интервал не интерактивной загрузки он регулировался с использованием манипулятора сети. T_3 и T_4 – аналогичные интервалы для интерактивной части эксперимента. Сумма $T_1+T_2+T_3+T_4$ представляет собой время сеанса связи.

Рисунок А.1/G.1030 – Интервалы времени в ходе эксперимента

Из-за известной разницы в поведении квалифицированных экспертов и так называемых неопытных, пользователей, при разработке модели между этими группами было сделано различие. Разделение этих групп позволяет нам разработать модель, которая прогнозирует качество поиска и просмотра информации в веб для большого числа пользователей.

А.3 Субъективные эксперименты по оценке качества поиска и просмотра информации в веб и их результаты

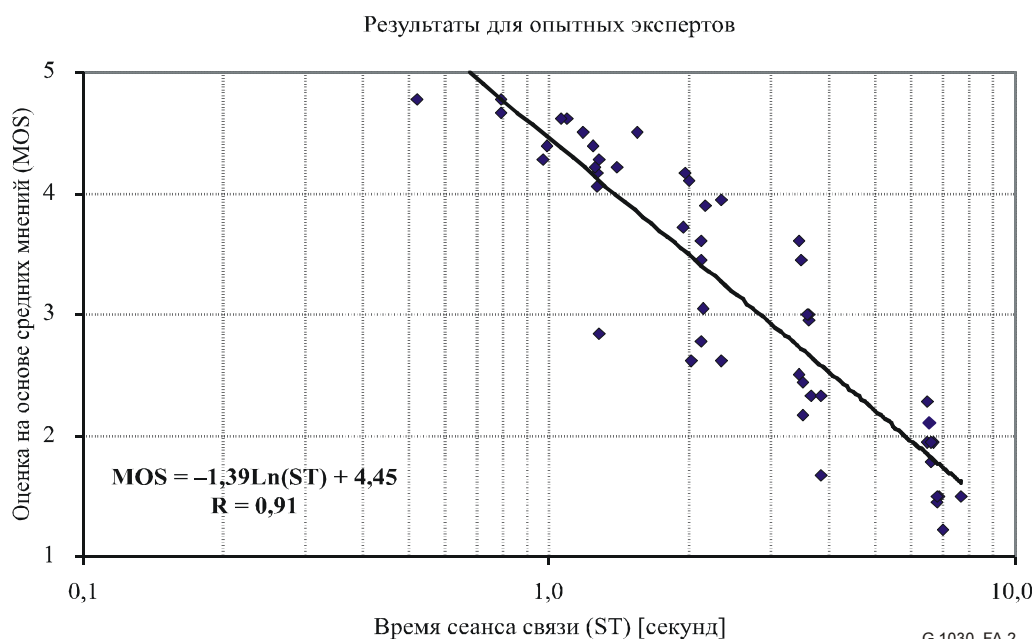
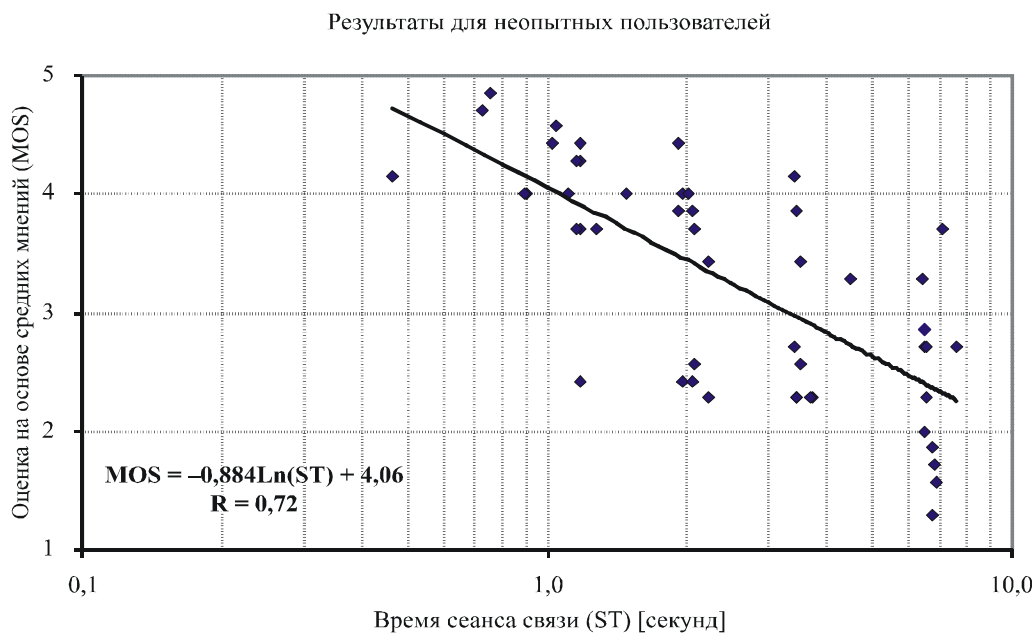
В экспериментах и, следовательно, в модели использована шкала абсолютных оценок МСЭ-Т [5] (пятибалльная шкала, в которой 5: отлично, 4: хорошо, 3: удовлетворительно, 2: плохо, 1: очень плохо). В каждом эксперименте было проведено 49 сеансов связи, каждый из которых состоял из:

- Запроса, получения и отображения страницы поиска.
- Набора и отправки предмета поиска на этой странице.
- Получения и отображения страницы результатов.

Для того чтобы получить непротиворечивые данные, в каждом сеансе связи использованы одинаковые страницы результатов, а пользователей просили в каждом сеансе связи печатать одинаковые запросы на поиск. С точки зрения пользователей, поисковая машина сначала должна была найти страницу результатов, которую затем она должна была загрузить. Для каждого из 49 сеансов связи, использовались различные комбинации интервалов T_1 – T_4 , составляющие различную сумму $T_1+T_2+T_3+T_4$, т. е., время сеанса связи, от 0 максимального времени для данного блока экспериментов.

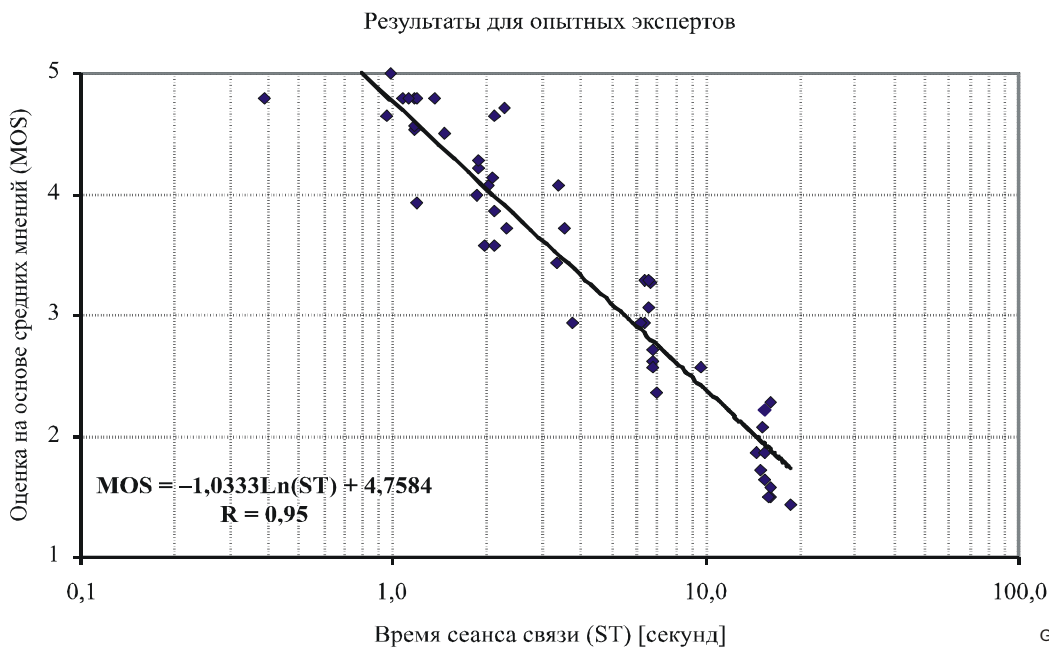
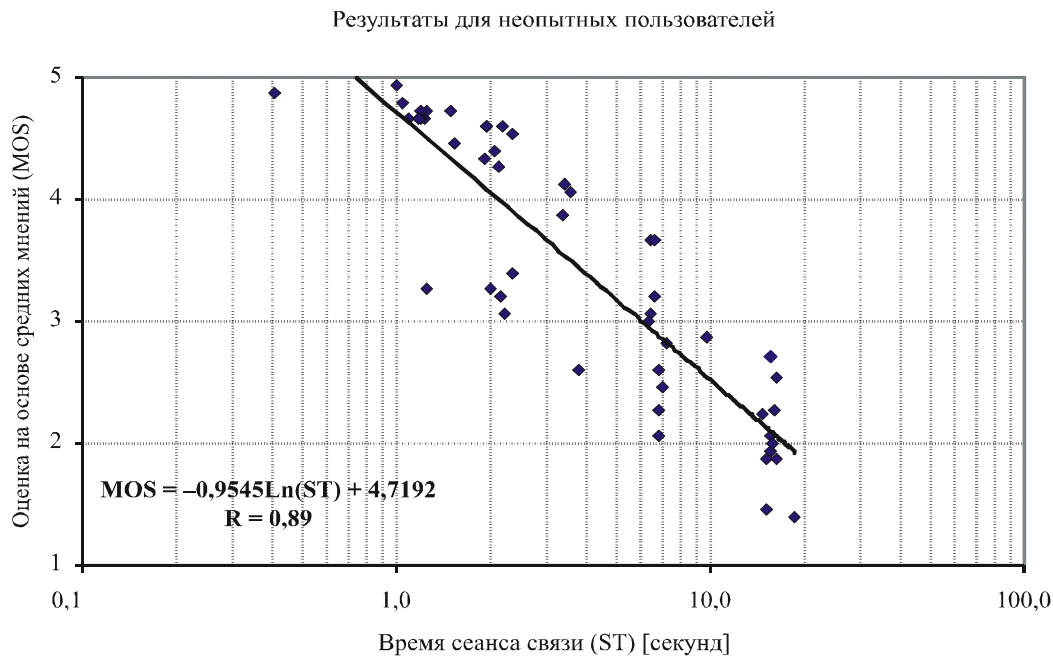
Общие результаты показаны на рисунках А.2–А.4, где изображено время сеанса связи (т. е. $T_1+T_2+T_3+T_4$) относительно значения MOS для всех трех экспериментов. Для продолжительной работы (см. рисунок А.4), результаты для неопытных пользователей и экспертов оказались примерно одинаковыми, и значения MOS были рассчитаны для всех пользователей. Для двух более коротких интервалов (см. рисунки А.2 и А.3), неопытные пользователи и эксперты вели себя различно, и корреляции между интервалами времени сеанса связи и воспринимаемым качеством работы оказались очень разными, и, поэтому, приведены по отдельности. Все результаты имеют схожий вид, воспринимаемое качество работы линейно уменьшается относительно логарифма от времени сеанса связи. Корреляция при продолжительных условиях эксперимента довольно велика ($>0,9$) для того,

чтобы надежно предсказать качество работы и для неопытных пользователей, и для экспертов. Как правило, в психологическом моделировании восприятия качества работы требуется корреляция более 0,9 [6]. Результаты показывают также, что в 6-секундном эксперименте с неопытными пользователями корреляция между временем сеанса связи и воспринимаемое качество работы очень мала (0,72) для того, чтобы сформировать простую модель, основанную только на времени сеанса связи.



Результаты для (7) неопытных пользователей и (18) экспертов с максимальным временем 6 секунд. Для неопытных пользователей корреляция слишком мала для правильного предсказания MOS. Для экспертов MOS можно спрогнозировать в виде логарифмической интерполяции времени сеанса связи от 0,67 до 12 секунд.

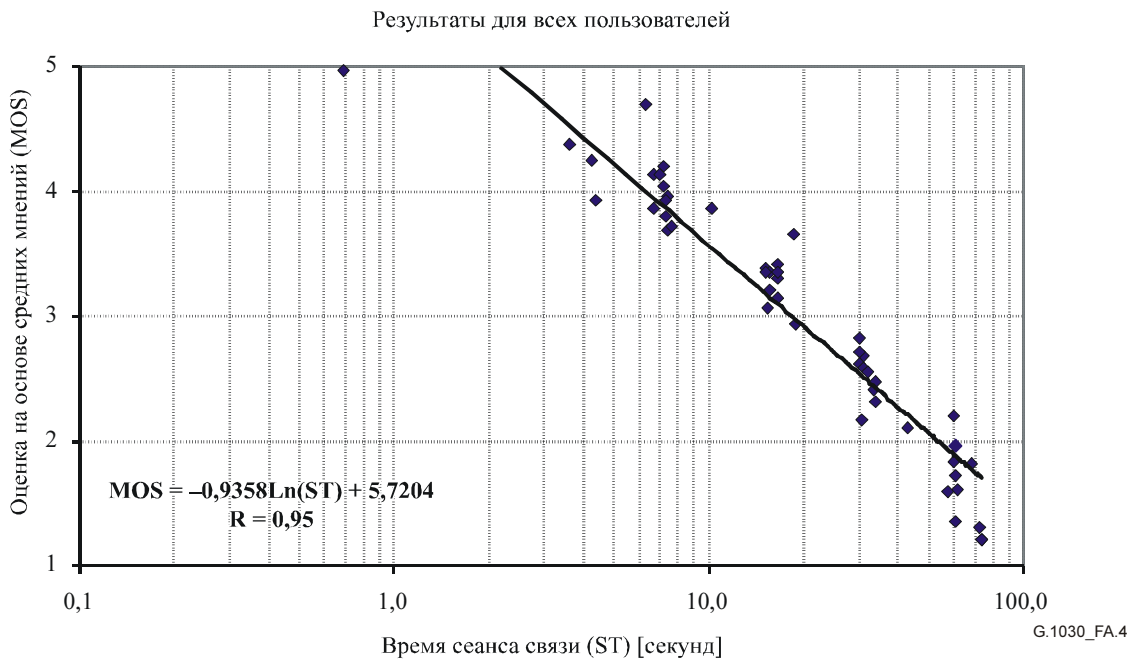
Рисунок А.2/G.1030 – Результаты субъективных испытаний для 6 секунд



G.1030_FA.3

Результаты для (15) неопытных пользователей и (14) экспертов с максимальным временем 15 секунд. Оба графика ведут себя одинаково. Для неопытных пользователей корреляция немного маловата для точных предсказаний MOS. Для экспертов значение MOS можно предсказать на основе логарифмической интерполяции времени сеанса связи от 0,79 до 38 секунд.

Рисунок А.3/Г.1030 – Результаты субъективных испытаний для 15 секунд



Результаты для всех (12+17=29) пользователей для времени 60 секунд. Значение MOS можно предсказать на основе логарифмической интерполяции времени сеанса связи от 2,16 до 155 секунд.

Рисунок А.4/Г.1030 – Результаты субъективных испытаний для 60 секунд

А.4 Результаты моделирования

А.4.1 60-секундные испытания

Для 60-секундных испытаний, корреляция времени сеанса связи и субъективного качества работы очень хорошая (0,95) и для экспертов, и для неопытных пользователей, и в объективной модели может непосредственно использоваться соответствие между временем сеанса связи и субъективным качеством. Регрессия показывает, что время сеанса связи менее 2 секунд приводит к получению максимального значения MOS = 5 (см. рис. А.4), тогда как минимальное значение MOS = 1 получается для сеансов связи длительностью более чем наиболее продолжительный сеанс связи в эксперименте. Общее соответствие между временем сеанса связи и качеством поиска и просмотра информации в веб для продолжительной работы формируется путем определения минимального (*Min*) и максимального (*Max*) сеанса связи и применения логарифмической интерполяции между этими крайними значениями времени. Если записать $MOS = a - b \cdot \ln(\text{Время Сеанса Связи})$, подставить $MOS = 5$ для *времени сеанса связи* = *Min* и $MOS = 1$ для *времени сеанса связи* = *Max*, получим значения сеанса связи между *Min* и *Max*:

$$MOS = \frac{4}{\ln(\text{Min} / \text{Max})} \cdot (\ln(\text{Время сеанса связи}) - \ln(\text{Min})) + 5. \quad (1)$$

Для продолжительного эксперимента (см. рисунок А.4) регрессия имеет вид:

$$MOS = 5,72 - 0,936 \cdot \ln(\text{Время сеанса связи}), \quad (2)$$

поделенные между MOS 1,0 и 5,0.

А.4.2 6-секундные и 15-секундные испытания

В 6-секундных и 15-секундных испытаний корреляция между временем сеанса связи и субъективным качеством работы намного меньше, чем для 60-секундных испытаний, в результате была сформирована более продвинутая модель предсказания субъективного качества работы для неопытных пользователей и экспертов, используя ту идею, что для более коротких сеансов связи последний интервал времени загрузки (в нашем эксперименте T_4) более серьезно влияет на окончательное воспринимаемое качество поиска и просмотра информации в веб, чем остальные интервалы времени отклика и загрузки (в нашем эксперименте T_1, T_2, T_3). В таблице А.1 приводятся весовые коэффициенты, с которыми должны быть взвешены значения T_1 – T_4 , для того

чтобы получить количество, которое максимально коррелируется с субъективно определенными значениями MOS. Это значение, взвешенное время сеанса связи,

$$\text{Взвешенное время сеанса связи} = WT1 \cdot T1 + WT2 \cdot T2 + WT3 \cdot T3 + WT4 \cdot T4.$$

Может быть сопоставлено со значением MOS, используя ту же самую логарифмическую интерполяцию между минимальным и максимальным временем сеанса связи, что использовано в уравнении (1):

$$MOS = \frac{4}{\ln(\text{Min} / \text{Max})} \cdot (\ln(\text{Взвешенное время сеанса связи}) - \ln(\text{Min})) + 5. \quad (3)$$

Из таблицы А.1 видно, что для самой короткой работы влияние последнего интервалы загрузки вдвое выше, чем влияние других интервалов отклика и загрузки. Из таблицы видно также разное поведение для неопытных пользователей и экспертов, оптимальное взвешивание демонстрирует более заметное влияние последнего интервала загрузки на мнение неопытных пользователей, чем на мнение экспертов. Для неопытных пользователей влияние последнего интервала времени загрузки более, чем вчетверо превосходит влияние других интервалов отклика и загрузки. Для экспертов это соотношение сводится к примерно двум, тогда как общее наилучшее взвешивание показывает, что влияние последнего интервала времени загрузки больше примерно втрое.

Таблица А.1/G.1030 – Оптимальное взвешивание в модели для интервалов времени T₁, T₂, T₃ и T₄ с соответствующими значениями корреляции в модели между объективными значениями времени и субъективными результатами MOS

	WT1	WT2	WT3	WT4	Min	Max	Корреляция
6 с эксперт	0,56	0,84	0,80	1,80			0,97
6 с неопытный пользователь	0,37	0,40	0,60	2,63			0,93
6 с общее значение	0,47	0,60	0,71	2,22	0,62	13,5	0,95
15 с эксперт	0,63	0,77	1,11	1,49			0,98
15 с неопытный пользователь	0,48	0,70	0,88	1,95			0,96
15 с общее значение	0,54	0,72	0,98	1,76	0,81	39	0,97
60 с эксперт	0,84	0,77	1,22	1,18			0,99
60 с неопытный пользователь	0,64	1,01	1,12	1,24			0,98
60 с общее значение	0,73	0,90	1,16	1,22	2,22	151	0,98
60 с общее значение, без взвешивания, см. рис А.4	1,00	1,00	1,00	1,00	2,16	155	0,95
<p>ПРИМЕЧАНИЕ. – Сумма весовых коэффициентов делится на 4,0 для того, чтобы иметь возможность сравнивать значения (T₁+T₂+T₃+T₄) для нормального сеанса связи со значениями для взвешенного сеанса связи. Взвешивание, используемое в стандартной модели вместе с минимальным и максимальным временем, используемым в уравнении (3), показано жирным шрифтом. Для продолжительных сеансов связи (>60 секунд) может быть использована простая модель без взвешивания (жирный курсив).</p>							

Из таблицы А.1 видно, что при средней продолжительности весовые коэффициенты для неопытных пользователей и экспертов, а также суммарные весовые коэффициенты лежат между весовыми коэффициентами для экспериментов малой и большой продолжительности. Это подтверждает справедливость применения взвешивания, и, таким образом, разрешает интерполяцию между различными значениями продолжительности экспериментальных сеансов связи (т. е. длительностью 6, 15 и 60 секунд) с целью получения весовых коэффициентов для сеансов связи другой длительности.

А.4.3 Резюме

В настоящей Рекомендации стандартизована простая модель для времени работы 60 секунд, которая с использованием уравнения (2) позволяет оценить сеансы связи поиска и просмотра информации в веб, для которых максимальное время сеанса связи составляет примерно 155 секунд. Кроме того, в

настоящей Рекомендации стандартизованы три усложненных модели, в которых используется наилучшие суммарные веса из таблицы А.1 вместе со следующими соотношениями взвешенного времени сеансов связи с воспринимаемым качеством поиска и просмотра информации в веб в единицах оценки среднего мнения:

$$MOS = 4,38 - 1,30 \cdot \ln(\text{Взвешенное время сеанса связи}), \quad (4)$$

поделенные между MOS 1,0 и 5,0 для коротких сеансов связи.

$$MOS = 4,79 - 1,03 \cdot \ln(\text{Взвешенное время сеанса связи}), \quad (5)$$

поделенные между MOS 1,0 и 5,0 для сеансов связи средней продолжительности.

$$MOS = 5,76 - 0,948 \cdot \ln(\text{Взвешенное время сеанса связи}), \quad (6)$$

поделенные между MOS 1,0 и 5,0 для длительных сеансов связи.

Результаты регрессии, которые соответствуют уравнениям (4), (5) и (6) показаны на рисунках А.5–А.7.

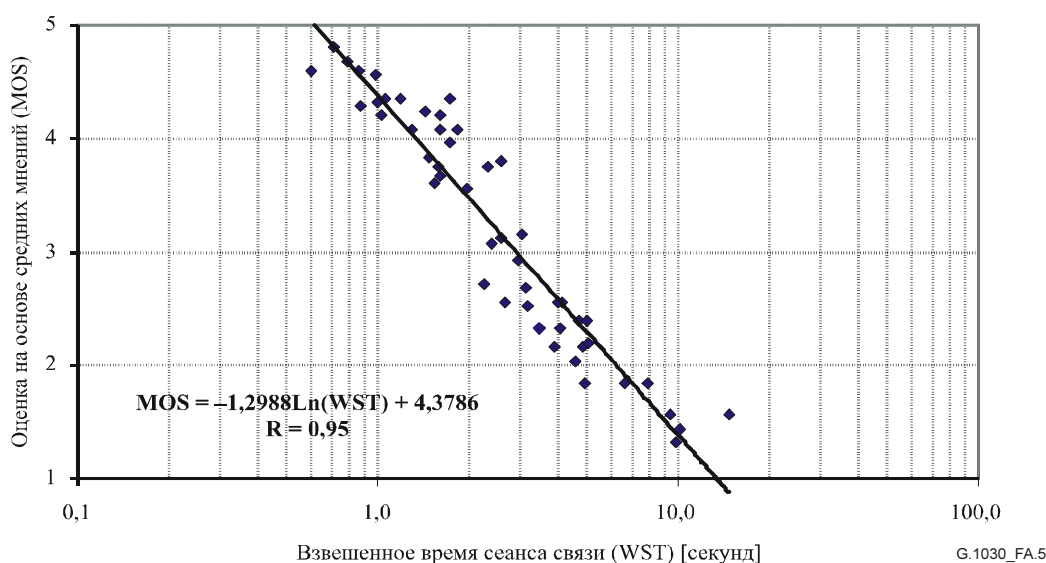


Рисунок А.5/G.1030 – Модель в зависимости от данных для всех пользователей для продолжительности 6 секунд

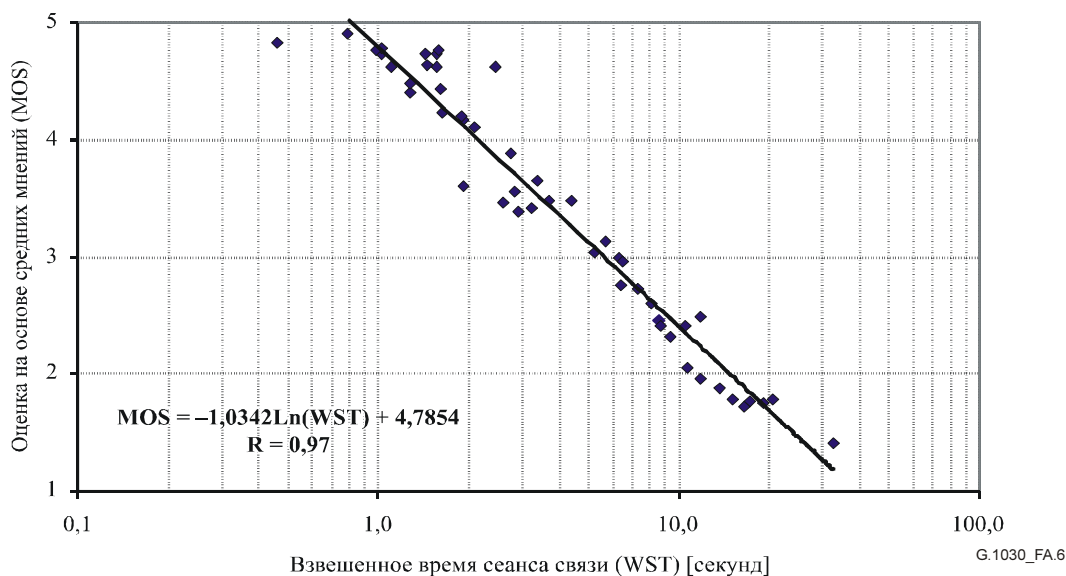


Рисунок А.6/G.1030 – Модель в зависимости от данных для всех пользователей для продолжительности 15 секунд

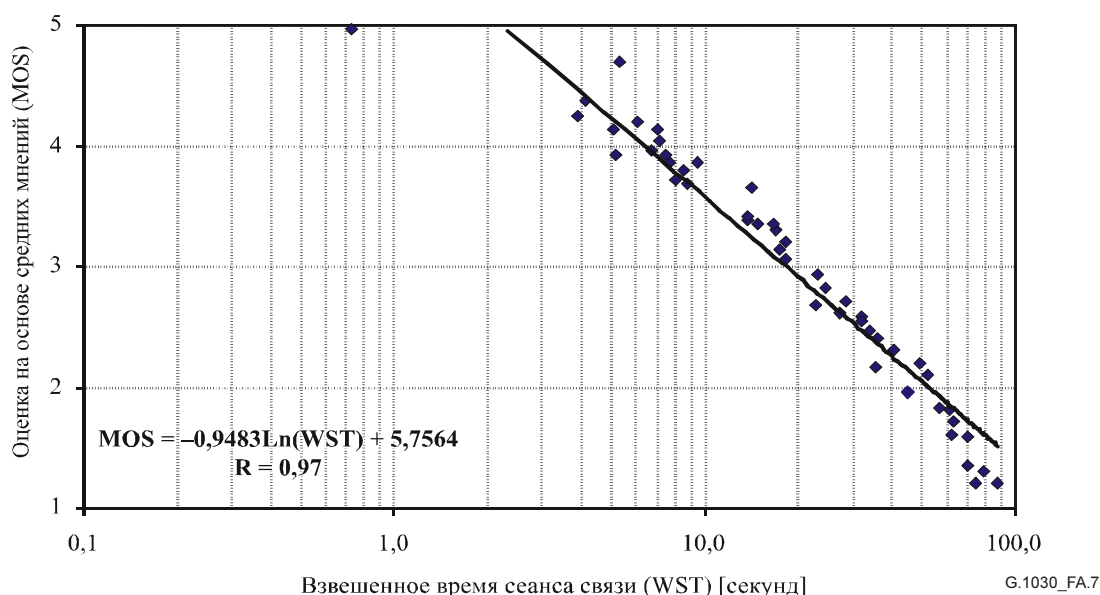


Рисунок А.7/G.1030 – Модель в зависимости от данных для всех пользователей для продолжительности 60 секунд

А.5 Воспринимаемое качество сеансов связи с просмотром одной веб-страницы и единичных событий (не стандарт)

На основании данных экспериментов, которые были начальной точкой для вывода уравнений (1) и (3), мы также вывели соотношение между временем поиска и просмотра информации в веб, состоящих из просмотра одной страницы, и воспринимаемым качеством работы. Кроме того, мы также рассматривали сеансы связи с просмотром одной страницы, для которых время загрузки всегда равно нулю. Такая ситуация возникает, когда запрашиваемая информация появляется перед пользователем мгновенно после некоторого времени ожидания. Это последнее соотношение описывает взаимосвязь единичного события с субъективно воспринимаемым качеством работы в единицах оценок MOS.

Вывод соответствия для одностраничного сеанса связи начинается с того наблюдения, что качество работы уменьшается линейно пропорционально логарифму времени сеанса связи между минимальным и максимальным временем сеанса связи, для которых оценки MOS составляют соответственно 5,0 и 1,0 (см. рисунки А.2–А.7). Из данных для двухстраничных сеансов связи поиска и просмотра информации в веб, которые показаны на рисунках А.2–А.4, видим, что минимальное время сеанса связи в трех рассмотренных экспериментальных поисках, меняется от 0,7 до 2,2, и увеличивается с ростом максимальной продолжительности эксперимента (см. обзор результатов в таблице А.2).

Таблица А.2/G.1030 – Невзвешенные минимальное и максимальное значения времени для двухстраничного эксперимента поиска и просмотра информации в веб

	<i>Min</i> (с)	<i>Max</i> (с)
6 с эксперт	0,67	12
15 с эксперт	0,79	38
60 с общее значение	2,16	155

Из этих результатов мы можем определить минимальное время сеанса связи (*Min*), для которого достигается оценка MOS = 5,0, как $Min = 0,011Max + 0,47$, где *Max* представляет собой максимальное ожидаемое время сеанса связи. Это позволяет получить общее соответствие времени сеанса связи значению оценки MOS для двухстраничного эксперимента при любой ожидаемой максимальной продолжительности (*Max*) двухстраничного сеанса связи поиска и просмотра информации в веб:

$$MOS_{2-стр} = \frac{4}{\ln((0,011Max+0,47)/Max)} \cdot (\ln(Время сеанса связи) - \ln(0,011Max+0,47)) + 5 \cdot \quad (7)$$

Из представленных экспериментальных данных, можно оценить, что это соотношение, вероятно, сохранится для всех двухстраничных сеансов связи поиска и просмотра информации в веб длительностью от 10 до примерно 200 секунд.

Для произвольных одностраничных сеансов связи, для которых имеется только один интервал времени отклика T_1 и один интервал времени загрузки T_2 (см. рисунок А.1) наивысшее качество получается для минимального значения, которое составляет примерно половину значения, полученного для двухстраничных сеансов связи. Это позволяет сформулировать определение минимального времени сеанс связи $Min = 0,005Max + 0,24$, при этом соотношение со значениями MOS описывается уравнением (8):

$$MOS_{1-стр} = \frac{4}{\ln((0,005Max+0,24)/Max)} \cdot (\ln(Время сеанса связи) - \ln(0,005Max+0,24)) + 5 \cdot \quad (8)$$

Из представленных экспериментальных данных, можно оценить, что это соотношение, вероятно, сохранится для всех одностраничных сеансов связи поиска и просмотра информации в веб сеансов связи длительностью от 5 до 100 секунд.

Аналогично, для произвольных одностраничных сеансов связи, для которых время загрузки всегда равно 0, или для любого единичного события, мы получаем:

$$MOS_{единичное событие} = \frac{4}{\ln((0,003Max+0,12)/Max)} \cdot (\ln(Время сеанса связи) - \ln(0,003Max+0,12)) + 5 \cdot \quad (9)$$

Из представленных экспериментальных данных, можно оценить, что это соотношение, вероятно, сохранится для всех единичных событий длительностью от 3 до 50 секунд. Отметим, что для такой одиночной процедуры минимальное время составляет 0,12 секунд, что соответствует порогу мгновенного восприятия [3].

Следует отметить, что описанные выше модели воспринимаемого качества одностраничных сеансов связи поиска и просмотра информации в веб и одиночных событий должны быть проверены реальными экспериментами.

А.6 Справочная литература к Приложению А

- [1] BEERENDS (J.G.), VAN DER GAAST (S.), AHMED (O.K.), Web browse quality modelling, *White contribution COM 12-C 3 to ITU-T Study Group 12*, November 2004.
- [2] VAN DER GAAST (S.), BEERENDS (J.G.), AHMED (O.K.), and MEEUWISSEN (H.B.), Quantification and prediction of end-user perceived web-browsing quality, submitted on March 24, 2005.
- [3] NIELSEN (J.), Response Times: The Three Important Limits (1994). Available: <http://www.useit.com/papers/responsetime.html>
- [4] DELLAERT (G.C.), KAHN (B.E.), How Tolerable is Delay? Consumers' Evaluations of Internet Websites after Waiting (1998). Available: <http://greywww.kub.nl:2080/greyfiles/center/1998/64.html>
- [5] ITU-T Recommendation P.800 (1996), *Methods for subjective determination of transmission quality*.
- [6] ITU-T Recommendation P.862 (2001), *Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs*.

Дополнение I

Оценки качества работы сети в условиях ограниченной информации

I.1 Введение

В настоящем дополнении содержится информация об упрощенных методах, которые могут использоваться для оценки сквозного качества работы в IP-сети, для которой имеется ограниченная информация. Точность оценок, полученных с использованием этих методов, сильно зависит от качества исходных данных. Материал данного дополнения будет меняться в зависимости от последующих исследований и оценок.

I.2 Эталонное соединение

Необходимо учитывать наличие систем система защиты доступа, прокси-серверов (для веб и/или усиления результативности), трансляторов сетевых адресов (NAT) коммутаторов распределения загрузки в эталонном соединении, поскольку эти устройства сегодня установлены в различных домах, учреждениях и управляемых IP-сетях. Эти "промежуточные блоки" выполняют различные функции, но разрушают сквозную прозрачность, которая является желательным элементом архитектуры IP-сети. На рисунке I.1 показано соответствующее эталонное соединение.

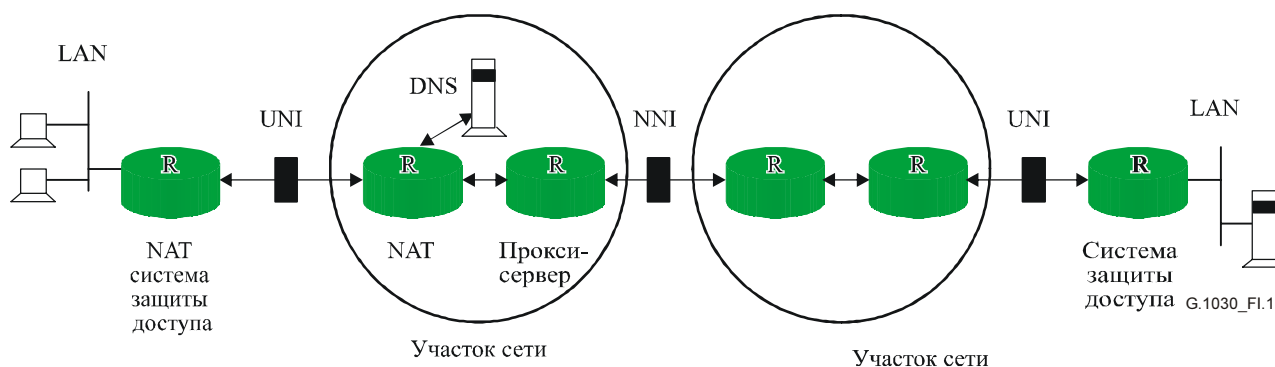


Рисунок I.1/G.1030 – Эталонное соединение с примерными промежуточными звеньями

Такое расширение позволяет оценить качественные показатели протоколов сигнализации, например SIP, используя те же самые основные принципы, что и в других IP-приложениях.

I.3 Комбинация значений оценок качества пакетной передачи

Введение понятия "промежуточные блоки" означает, что в сквозной линии связи будут существовать отдельные участки, что определяет необходимость нахождения более точных комбинационных формул.

I.3.1 Задержка

Средние значения задержки для отдельных участков суммируются. Однако можно отметить, что средняя задержка представляет собой центр тяжести распределения изменения задержки, которая часто наблюдается при пакетной передаче. Когда задержка передачи пакетов на участке сети (или время обработки на хосте) представлено средним значением, отсчет распределения задержки заменяется его ожидаемым значением. Здесь вполне можно использовать средние значения времени задержки, поскольку каждая транзакция будет анализировать соответствующее распределение задержки во множестве экземпляров во время транзакции.

I.3.2 Потери

Показатель коэффициента сквозных потерь (IPLR) является чрезвычайно важным для качества UNI-UNI. Можно предположить, что вероятности потерь на конечных терминалах и хостах ничтожны.

Метод, предлагаемый для объединения значений IPLR, состоит в том, чтобы инвертировать вероятность успешной передачи пакета на протяжении n участков сети, следующим образом:

$$IPLR_{UNI-UNI} = 1 - \{(1 - IPLR_{NS1}) \times (1 - IPLR_{NS2}) \times (1 - IPLR_{NS3}) \times \dots \times (1 - IPLR_{NSn})\}.$$

В основе этого уравнения лежит теория условной вероятности, оно предполагает также, что вероятности потерь на каждом участке сети не зависят друг от друга. Для трассы UNI-UNI, состоящей из двух сетей А и В с вероятностями потерь p_A и p_B :

$\text{Prob}\{\text{успешная передача в обеих сетях}\} = \text{Prob}\{\text{успешная передача в сети В} | \text{успешная передача в сети А}\} \times \text{Prob}\{\text{успешная передача в сети А}\} = (1 - p_B) \times (1 - p_A).$

Эмпирически полученные вероятности потерь для отдельных сетей (p_A и p_B) эквивалентны условным вероятностям успешной передачи в любой предшествующей сети – для начала операций измерения требуется получение пакета.

1.3.2.1 Суммирование импульсных потерь пакетов

Требует дальнейшего изучения.

1.3.3 Изменение времени задержки

Традиционные IP-приложения и многие приложения передачи данных менее чувствительны к изменению времени задержки, чем приложения, требующие непрерывного воспроизведения (изохронные приложения или приложения реального времени, например, VoIP). По этой причине традиционные IP-приложения иногда классифицируются по "эластичности" пакетных потоков, которые во время передачи могут сжиматься или растягиваться, почти не оказывая влияния на результат. Когда при транспортировке VoIP меняется расстояние между пакетами, для приложения требуется дополнительная буферизация для восстановления возможности непрерывного воспроизведения, но для "неэластичных" пакетных потоков сам буфер является источником задержки.

Довольно трудно корректным образом объединить изменение задержки на различных участках сети и в различных устройствах обработки, поскольку полное распределение задержки редко известно. Однако, в разделе 8 Рекомендации МСЭ-Т Y.1541 приводится разумный метод аппроксимации.

1.4 Полоса пропускания "узкого места"

Использование доступа к полосе пропускания в качестве лимитирующего фактора при расчете времени сквозной передачи данных можно получить чрезмерно оптимистичную оценку пропускной способности.

1.4.1 Ключевые параметры

Более точная аппроксимация реального времени передачи файла фиксированного размера или сообщения может быть выполнена путем учета в модели следующих соображений:

- 1) Должно быть известно направление передачи. Для многих новых технологий, таких как DSL кабельные модемы скорости доступа асимметричны, и степень различия может достигать десяти и более.
- 2) Используйте реальную ширину полосы пропускания критического параметра. Критическим параметром не обязательно будет скорость доступа. Один из хостов может снижать скорость передачи (если ему не удастся передать достаточно данных в процессы TCP, или при наличии ограничений окна), параметры хоста TCP могут ограничить пропускную способность, и промежуточные блоки, такие как повышающие качество прокси-серверы могут улучшать параметры TCP для повышения пропускной способности, и т. д.
- 3) В ходе транзакции может встретиться соперничество за совместно используемые ресурсы. Используйте эффективную полосу пропускания критического параметра, определенную проектными требованиями, для поддержания определенного числа одновременных пользователей в одной линии или в одном совместно используемом ресурсе.

- 4) Полоса пропускания критического параметра может изменяться во времени. Ограничения полосы пропускания приведут к получению широкого диапазона значений времени передачи.
- 5) Дополнительная нагрузка должна рассчитываться в точке критического параметра. Каждый уровень ниже IP добавляет заголовки со служебными данными. Предположите объем нагрузки и рассчитайте процент, занимаемый заголовками, он еще больше увеличивает время передачи данных.
- 6) TCP не достигает пропускной способности, равной ширине полосы пропускания. Уравнение Матиса для TCP является хорошей аппроксимацией пропускной способности передачи в установившемся режиме, хотя вышеприведенные факторы могут иметь значение, превышающее граничное.

Большая часть этих пунктов требует лучшей подготовки со стороны того, кто использует модель или того, кто планирует сеть. Подробности аппроксимации пропускной способности TCP приводятся в п. I.4.2.

I.4.2 Пропускная способность TCP

Преобладающее большинство традиционных IP-приложений используют надежные службы передачи байтовых потоков TCP (Протокола управления передачей). Модель управления потоком TCP описывается следующим образом [TCPCon]:

$$BW < \frac{MSS}{RTT} \frac{C}{\sqrt{p}},$$

где:

BW: данные, переданные в единицу времени (время цикла)

MSS: максимальный размер сегмента TCP

RTT: среднее время передачи в оба конца

C: постоянная, которая учитывает влияние случайных/периодических потерь и стратегии АСК

p: вероятность потери пакетов.

Это уравнение следует рассматривать как верхнюю границу информационного потока, даже для самых последних модификаций TCP, таких как SACK и TCP Reno. В нем предполагается, что соединение TCP длится достаточно долго для достижения равновесия в состоянии предотвращения перегрузки. Когда общее число байтов в передаваемых данных мало, управление потоком TCP может не достигать равновесия, и в расчетах доминирует поведение медленного старта. Более гибкое соотношение для пропускной способности TCP Reno можно найти в [Padhye], оно учитывает лимитирующее влияние максимального размера окна.

$$B(p) \approx \min \left(\frac{W_{\max}}{RTT}, \frac{1}{RTT \sqrt{\frac{2bp}{3}} + T_0 \min \left(1, 3\sqrt{\frac{3bp}{8}} \right) p (1 + 32p^2)} \right),$$

где:

B(p): примерная модель пропускной способности TCP [пакеты/с]

W_{max}: максимальный размер буфера окна приемника [пакеты]

RTT: время двусторонней передачи [с]

b: число пакетов, прием которых подтвержден принятым АСК

p: вероятность того, что пакет потерян

T₀: время простоя для повторной передачи неподтвержденного (потерянного) пакета [с].

Управление потоком TCP стремится увеличивать скорость передачи (размер окна) до тех пор, пока не возникает перегрузка (или "узкое место" в линии), и оно делает вывод, что это происходит из-за потери пакетов. Таким образом, некоторая степень потери пакетов свойственна проверка с

использованием критического параметра. Способ предотвращения перегрузки состоит в управлении потоком, которое при возникновении потерь уменьшает размер окна вполнину, а затем увеличивает окно на один пакет за раз до тех пор, пока не будет обнаружена новая потеря. Этот способ эффективно ограничивает пропускную способность до значения, составляющего примерно 75% от пикового размера окна или скорости передачи. Это уравнение можно переписать, сохранив в нем только потери пакетов из-за TCP проверки с использованием критического параметра.

$$\sqrt{P_{\text{Проверка}}} \approx \frac{MSS}{RTT} \frac{C}{0,75 \times BW(\text{узкое_место_линии})}$$

В таблице I.1 перечислены присущие потери пакетов, связанные с конкретной скоростью в "узком месте" на линии.

Таблица I.1/G.1030 – TCP потери пакетов из-за проверки с использованием критического параметра

C	MSS	RTT	BW (линия)	75% BW	Задержка*BW	Окно	p
0,866	12 000	0,08	10 000 000	7 500 000	800 000	66,666667	3,00E-04
0,866	12 000	0,08	1 536 000	1 152 000	122 880	10,24	1,27E-02
0,866	12 000	0,08	768 000	576 000	61 440	5,12	5,09E-02
0,866	12 000	0,08	384 000	288 000	30 720	2,56	2,03E-01
0,866	12 000	0,08	128 000	96 000	10 240	0,8533333	1,83E+00
0,866	12 000	0,08	64 000	48 000	5120	0,4266667	7,32E+00
0,866	2048	0,08	128 000	96 000	10 240	5	5,33E-02
0,866	2048	0,08	64 000	48 000	5120	2,5	2,13E-01

По таблице I.1 можно сделать несколько замечаний:

- Присущие потери пакетов очень малы, когда BW "узкого места" на линии составляет 10 Мбит/с (Ethernet или выше). Если потери пакетов в сети 10^{-4} и выше, это может привести к снижению пропускной способности в соответствии с [TCPCon].
- Для данных параметров (80 мс RTT, 12 000 битов (1500 байтов) MSS, и т. д.) примерно один пакет из 100 теряется на интервале T_1 , когда один-единственный поток TCP проверяет "узкое место". Потери пакетов в сети с вероятностью 10^{-3} и ниже могут почти не сказываться на результирующей пропускной способности.
- Показаны произведение Задержка*BW и оптимальный размер (произведение/MSS) для трассы, отметим, что существует две ширины полосы (BW) линии (128 кбит/с и 64 кбит/с), когда окно меньше единицы и рассчитанный коэффициент потерь не имеет смысла (>1). В этих случаях гарантируется регулировка TCP, поэтому мы уменьшаем MSS до величины 2048 битов и получаем более разумный результат.
- Отметим, что здесь не рассматривается уменьшение BW из-за наличия заголовка, созданного на нижнем уровне.

Оцененный коэффициент потерь в сети будет комбинироваться с коэффициентом потерь при проверке следующим образом:

$$P_{\text{Сумма}} = 1 - \{(1 - P_{\text{Сеть}}) \times (1 - P_{\text{Проверка}})\}$$

Когда коэффициенты потери пакетов в сети и при проверке имеют один и тот же порядок величин, немного пессимистично рассчитывать BW, комбинируя их, потому что некоторые из потерь в проверке даже и не возникнут.

И последнее, можно отметить, здесь совершенно не рассматривается регулировка параметров TCP, кроме того, что эти параметры могут каким-либо образом ограничивать пропускную способность. Если максимальный размер окна недостаточен для заполнения пути двусторонней передачи, то пропускная способность будет ограничена значением Размер окна \times MSS/RTT [Padhye]. Во многих TCP соединениях никогда не бывает потерь пакетов на протяжении всего их существования, благодаря малым размерам окна и стремлению TCP выстроить очередь пакетов переда "узком

местом" (и увеличению RTT для некоторых пакетов – именно поэтому используется среднее значение RTT).

I.5 Время соединения

Обмен пакетами можно разделить на две фазы:

- 1) Соединение; и
- 2) Передача данных (в некоторых приложениях этой фазы нет, например, в терминалах торговых точек для кредитных карт и сигнализация в VoIP. Для упрощения мы включили в эту фазу передачу финального пакета. Традиционная матрица 3×3 рассматривает рассоединение как отдельную фазу).

Если пакет или ответ на него теряется в ходе начальной установки соединения, передатчик, обычно, ожидает в течение определенного промежутка времени, прежде чем выполнить повторную передачу сообщения. Время ожидания, или время простоя из-за повторной передачи может быть от одной до трех секунд, и при возникновении потери существенно увеличивает первоначальное время соединения.

Если все интервалы времени, (включая те, что обусловлены потерями) отделены от интервалов времени обработки на хосте, существует два параметра, которые связывают разделяемые административные зоны ответственности (сеть и СРЕ/хосты).

Такое разделение довольно прямолинейно для начальной фазы установки соединения. Время передачи пакетов и время простоя из-за потерь относятся к сетям.

Для времени передачи данных может быть достаточным указать это время и отметить, что было лимитирующим фактором – потери в сети, задержка в сети или обработка/установка хоста.

I.5.1 Влияние потерь во время установки соединения

Вклад IP-сети в исходное время соединения можно определить, как изображено на рисунке ниже. Сначала рассмотрим случай, когда должен быть выполнен обмен 8 сообщений и ответов, и все передачи пакетов были успешными. Вероятность этого случая показана ниже:

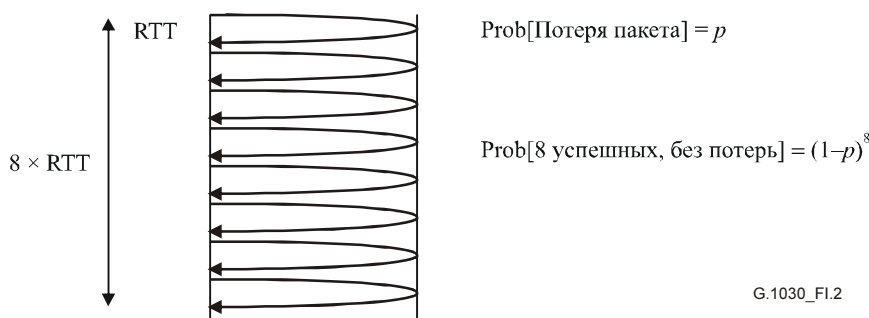


Рисунок I.2/G.1030 – Соединение с 8 повторами приема-передачи, без потери пакетов

Это – простой случай, но он представляет собой вклад IP-сети в то ощущение, которое складывается у пользователя, когда вероятность потери пакетов при двусторонней передаче (p) достаточно мала. Отметим, что вероятность потери при двусторонней передаче равна:

$$p = p_{RT} = 1 - \{(1 - p_{1-way}) \times (1 - p_{other-way})\}.$$

Время для выполнения 8 двусторонних передач в восемь раз превышает среднее время одной двусторонней передачи (RTT), плюс удаленная обработка на хосте (НРТ), и в результате получаем время соединения. Эти компоненты можно рассчитать по отдельности, как:

$$\text{Время_соединения} = NCTT + \text{Сумма_НРТ},$$

где NCTT – это вклад сети во время передачи, определенное в Рекомендации МСЭ-Т G.1040.

Отметим, что пример соединения РОР3 содержит периодов обмена 8 запрос/ответ. Если предположить, что $p = 10^{-3}$, $RTT + НРТ = 0,080$ секунд, время просто из-за повторной передачи (RTO)

= 1 секунду, и имеется 350 тысяч попыток передачи, получим значения вероятности для каждого из случаев потерь, показанных в таблице I.2.

Таблица I.2/G.1030 – Время соединения для случаев с потерей от 0 до 3 пакетов

Потери	Время соединения [с]	Вероятность появления	Число транзакций
0	0,64	0,99202794	347 210
1	1,64	0,00793622	2778
2	2,64	3,5713E-05	12
3	3,64	1,1904E-07	0

В этом примере показано, что вклад IP-сети составляет в среднем менее 1 секунды, 95-й перцентиль и 99-й перцентиль времени соединения для транзакции POP3 с 8 периодами обмена прием-передача (RT).

I.5.2 Влияние потери пакетов во время передачи данных

В результате потери пакетов потребуются повторная передача, при этом по процедуре предупреждения перегрузки регулирование потока TCP уменьшает скорость передачи, как описано в п. I.4.

I.6 Пример приложения: транзакция HTTP

На рисунке I.3 показана транзакция HTTP, определяющая время соединения. Оставшаяся часть передачи пакетов образует время передачи данных с точки зрения клиента.

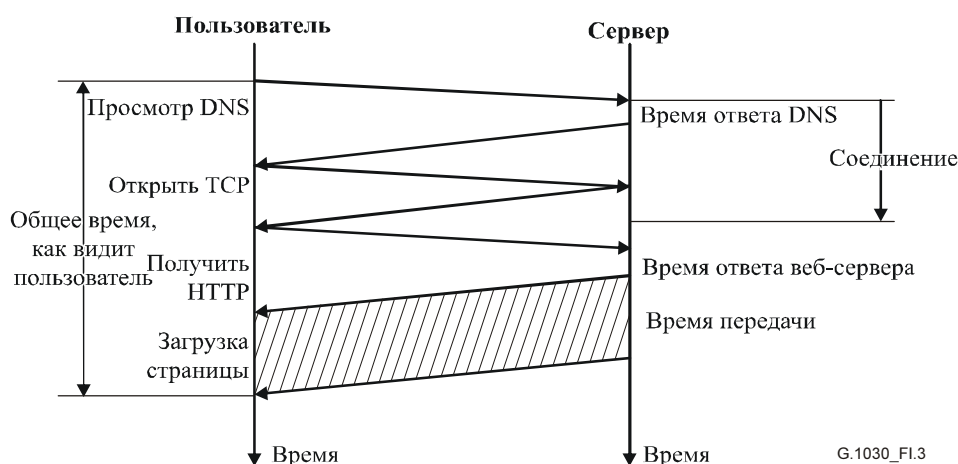


Рисунок I.3/G.1030 – Обмен по протоколу HTTP

Используется эталонное соединение, показанное на рисунке I.1, на нем показано два участка сети с NAT, системами защиты доступа и прокси-серверами в качестве примеров промежуточных блоков. Отметим, что прокси-сервер не содержит запрашиваемых объектов, сохраненных в кэш-памяти, а направляет запрос на удаленный веб-сервер.

I.6.1 Время соединения

Существует два запроса-ответа, передаваемых по различным трассам, DNS опрашивает, а TCP выполняет 3-стороннее соединение (в котором установлено расписание передачи для SYN и SYN-ACK, предполагается, что последнее подтверждение ACK обрабатывается вместе с командой ПОЛУЧИТЬ HTTP и не добавляет заметного времени). Хотя прокси-сервер разделяет TCP соединение, в этом примере мы рассматриваем его как дополнительную задержку в одной линии связи.

Поскольку большая часть операций комбинирования представляют собой простые сложение и умножение, мы просто свели значения в таблицу и показали результаты (вместо переписывания уравнений с множеством новых параметров).

Во-первых, в таблице I.3 показаны компоненты качества DNS.

Таблица I.3/G.1030 – DNS опрос

Клиент к DNS	RTT (чистое время) [мс]	НРТ (Время обработки) [мс]	RT (1-р)	BW на 1 поток
LAN	2		0,99999	5,0E+6
NAT/система защиты доступа		4	0,99999	
Линия 1	10		0,9999	5,0E+6
NAT/периферийный маршрутизатор			0,99999	
NAT к линии DNS	4		0,99999	10,0E+6
DNS		6	0,99999	
Сумма (инвертированные потери)	16	10	0,00015	

И, наконец, компоненты установки TCP соединения показаны в таблице I.4.

Таблица I.4/G.1030 – Трасса клиент-сервер для времени соединения TCP и времени передачи данных

Клиент к DNS	RTT (чистое время) [мс]	НРТ (Время обработки) [мс]	RT (1-р)	BW на 1 поток
LAN	2		0,99999	5,0E+6
NAT/система защиты доступа		4	0,99999	
UNI линия 1	10		0,9999	5,0E+6
NAT/периферийный маршрутизатор			0,99999	
NAT к линии прокси	4		0,99999	10,0E+6
Прокси-сервер		4	0,99999	
NNI линия	20		0,99999	10,0E+6
Участок сети 2	14		0,99999	10,0E+6
UNI линия 2	10		0,99999	3,0E+6
Система защиты доступа		2	0,99999	
LAN	2		0,99999	5,0E+6
Сервер (обработка TCP)		1	0,99999	
Сумма (инвертированные потери)	62	11	0,00021	

Если время просмотра DNS = 16 + 10 = 26 мс, а время открытия TCP 62 + 11 = 73 мс, то время соединения = 99 мс, и это время, как представляется, является правильным для более чем 99,9% транзакций HTTP.

1.6.2 Время передачи данных

Первые шаги должны определить, будут ли в данных расчетах играть заметную роль потери клиент-сервер.

Снова обращаясь к таблице I.4, мы видим, что наименьшая BW на 1 поток возникает в UNI линии 2 на скорости Мбит/с. Произведение Задержка × BW равно $0,073 \times 3M = 219$ кбит/с, или оптимальный размер окна составляет 18,25 пакетов на скорости 12 000 битов (1500 байтов) MSS. TCP можно отрегулировать, чтобы обеспечить такую трассу, таким образом, нет никаких ограничений, связанных с размером окна.

Используя BW "узкого места" линии, рассчитаем присущий TCP процент потери пакетов из-за проверок следующим образом:

$$\sqrt{p_{\text{Проверка}}} \approx \frac{MSS}{RTT} \frac{C}{0,75 \times BW}, \text{ таким образом } p = \left(\frac{12\,000}{0,073} \frac{0,866}{0,75 \times 3M} \right)^2 = 0,004003.$$

Поскольку мы рассчитали, что потери в сети составят (0,00021) и равны примерно $p_{\text{Проверка}}/20$, считаем их незначительными, тогда пропускная способность байтового потока TCP составит $0,75 \times 3 \text{ Мбит/с} = 2,25 \text{ Мбит/с}$.

Если, с другой стороны, оцененный коэффициент потерь в сети будет того же порядка, что и коэффициент потерь пакетов при проверке, то будет несколько пессимистично скомбинировать, а затем пересчитать BW:

$$p_{\text{Сумма}} = 1 - \{(1 - p_{\text{Сеть}}) \times (1 - p_{\text{Проверка}})\}$$

Предполагая работу с Веб-страницей со сложной графикой, будем считать, что время передачи для 1 Мбит составит в 0,444 секунды. Можно добавить время, необходимое серверу для обработки ПОЛУЧИТЬ HTTP (0,01 секунды), в результате получим 0,454 секунды.

I.6.3 Суммарное время, воспринимаемое пользователем

Итог – это сумма времени соединения и времени передачи данных $0,099 + 0,454 = 0,553$ секунд.

I.7 Резюме

В настоящем Дополнении описана методика оценки сквозного качества работы в IP-сетях. Полученные результаты могут использоваться для сравнения с поставленными задачами, например, Рек. МСЭ-Т G.1010, для оценки общего значения удовлетворенности конечного пользователя различными мультимедийными приложениями.

I.8 Справочная литература к Дополнению I

- ITU-T Recommendation G.1010 (2001), *End-user multimedia QoS categories*.
 - ITU-T Recommendation Y.1541 (2006), *Network performance objectives for IP-based services*.
- [Padhye] PADHYE (J.), FIROIU (V.), TOWSLEY (D.), and KUROSE (J.), Modeling TCP Throughput: a Simple Model and its Empirical Validation, SIGCOMM 1998.
<ftp://gaia.cs.umass.edu/pub/Padhye-Firoiu98:TCP-throughput.ps.Z>
- [TCPCon] MATHIS (M.), SEMKE (J.), MADAVI (J.), OTT (T.), The macroscopic behavior of TCP congestion avoidance algorithm, Computer communications review, Vol. 27, No. 3, July 1997, ISSN# 0146-4833.
http://www.psc.edu/networking/papers/model_ccr97.ps

СЕРИИ РЕКОМЕНДАЦИЙ МСЭ-Т

Серия А	Организация работы МСЭ-Т
Серия D	Общие принципы тарификации
Серия E	Общая эксплуатация сети, телефонная служба, функционирование служб и человеческие факторы
Серия F	Нетелефонные службы электросвязи
Серия G	Системы и среда передачи, цифровые системы и сети
Серия H	Аудиовизуальные и мультимедийные системы
Серия I	Цифровая сеть с интеграцией служб
Серия J	Кабельные сети и передача сигналов телевизионных и звуковых программ и других мультимедийных сигналов
Серия K	Защита от помех
Серия L	Конструкция, прокладка и защита кабелей и других элементов линейно-кабельных сооружений
Серия M	Управление электросвязью, включая СУЭ и техническое обслуживание сетей
Серия N	Техническое обслуживание: международные каналы передачи звуковых и телевизионных программ
Серия O	Требования к измерительной аппаратуре
Серия P	Качество телефонной передачи, телефонные установки, сети местных линий
Серия Q	Коммутация и сигнализация
Серия R	Телеграфная передача
Серия S	Оконечное оборудование для телеграфных служб
Серия T	Оконечное оборудование для телематических служб
Серия U	Телеграфная коммутация
Серия V	Передача данных по телефонной сети
Серия X	Сети передачи данных, взаимосвязь открытых систем и безопасность
Серия Y	Глобальная информационная инфраструктура, аспекты межсетевых протоколов и сети последующих поколений
Серия Z	Языки и общие аспекты программного обеспечения для систем электросвязи