

Unión Internacional de Telecomunicaciones

UIT-T

SECTOR DE NORMALIZACIÓN
DE LAS TELECOMUNICACIONES
DE LA UIT

G.1030

(11/2005)

SERIE G: SISTEMAS Y MEDIOS DE TRANSMISIÓN,
SISTEMAS Y REDES DIGITALES

Calidad de servicio y de transmisión – Aspectos genéricos
y aspectos relacionados al usuario

**Estimación de la calidad de funcionamiento
de extremo a extremo en redes IP para
aplicaciones de datos**

Recomendación UIT-T G.1030

RECOMENDACIONES UIT-T DE LA SERIE G
SISTEMAS Y MEDIOS DE TRANSMISIÓN, SISTEMAS Y REDES DIGITALES

CONEXIONES Y CIRCUITOS TELEFÓNICOS INTERNACIONALES	G.100–G.199
CARACTERÍSTICAS GENERALES COMUNES A TODOS LOS SISTEMAS ANALÓGICOS DE PORTADORAS	G.200–G.299
CARACTERÍSTICAS INDIVIDUALES DE LOS SISTEMAS TELEFÓNICOS INTERNACIONALES DE PORTADORAS EN LÍNEAS METÁLICAS	G.300–G.399
CARACTERÍSTICAS GENERALES DE LOS SISTEMAS TELEFÓNICOS INTERNACIONALES EN RADIOENLACES O POR SATELITE E INTERCONEXIÓN CON LOS SISTEMAS EN LÍNEAS METÁLICAS	G.400–G.449
COORDINACIÓN DE LA RADIOTELEFONÍA Y LA TELEFONÍA EN LÍNEA	G.450–G.499
CARACTERÍSTICAS DE LOS MEDIOS DE TRANSMISIÓN	G.600–G.699
EQUIPOS TERMINALES DIGITALES	G.700–G.799
REDES DIGITALES	G.800–G.899
SECCIONES DIGITALES Y SISTEMAS DIGITALES DE LÍNEA	G.900–G.999
CALIDAD DE SERVICIO Y DE TRANSMISIÓN – ASPECTOS GENÉRICOS Y ASPECTOS RELACIONADOS AL USUARIO	G.1000–G.1999
CARACTERÍSTICAS DE LOS MEDIOS DE TRANSMISIÓN	G.6000–G.6999
DATOS SOBRE CAPA DE TRANSPORTE – ASPECTOS GENÉRICOS	G.7000–G.7999
ASPECTOS RELATIVOS AL PROTOCOLO ETHERNET SOBRE LA CAPA DE TRANSPORTE	G.8000–G.8999
REDES DE ACCESO	G.9000–G.9999

Para más información, véase la Lista de Recomendaciones del UIT-T.

Recomendación UIT-T G.1030

Estimación de la calidad de funcionamiento de extremo a extremo en redes IP para aplicaciones de datos

Resumen

Esta Recomendación proporciona un marco de herramientas para obtener la calidad de funcionamiento de redes IP, estimar la calidad de funcionamiento de las aplicaciones de usuario y aplicar modelos de percepción para evaluar la satisfacción del usuario con la calidad de funcionamiento de extremo a extremo.

La calidad de funcionamiento que percibe el usuario de aplicaciones de datos en redes de paquetes depende de muchos factores, incluyendo la calidad de funcionamiento de extremo a extremo de la red de paquetes, la medida en que la aplicación depende de la red de comunicaciones, la calidad de funcionamiento de los terminales y otros dispositivos ajenos al ámbito de los operadores de red, así como las tareas del usuario y el grado de interacción del usuario con la aplicación. Los diseñadores de redes tienen estos factores en cuenta para asegurar la satisfacción del usuario. Una vez estimada la calidad de funcionamiento de la aplicación, pueden aplicarse modelos de percepción para interpretar el nivel de calidad de funcionamiento de extremo a extremo.

En esta Recomendación se supone que el lector es capaz de aportar un mínimo de detalles con respecto a cada uno de los factores clave anteriores, y de utilizar el marco de herramientas para poder estimar la calidad de funcionamiento de extremo a extremo.

Orígenes

La Recomendación UIT-T G.1030 fue aprobada el 29 de noviembre de 2005 por la Comisión de Estudio 12 (2005-2008) del UIT-T por el procedimiento de la Recomendación UIT-T A.8.

PREFACIO

La UIT (Unión Internacional de Telecomunicaciones) es el organismo especializado de las Naciones Unidas en el campo de las telecomunicaciones. El UIT-T (Sector de Normalización de las Telecomunicaciones de la UIT) es un órgano permanente de la UIT. Este órgano estudia los aspectos técnicos, de explotación y tarifarios y publica Recomendaciones sobre los mismos, con miras a la normalización de las telecomunicaciones en el plano mundial.

La Asamblea Mundial de Normalización de las Telecomunicaciones (AMNT), que se celebra cada cuatro años, establece los temas que han de estudiar las Comisiones de Estudio del UIT-T, que a su vez producen Recomendaciones sobre dichos temas.

La aprobación de Recomendaciones por los Miembros del UIT-T es el objeto del procedimiento establecido en la Resolución 1 de la AMNT.

En ciertos sectores de la tecnología de la información que corresponden a la esfera de competencia del UIT-T, se preparan las normas necesarias en colaboración con la ISO y la CEI.

NOTA

En esta Recomendación, la expresión "Administración" se utiliza para designar, en forma abreviada, tanto una administración de telecomunicaciones como una empresa de explotación reconocida de telecomunicaciones.

La observancia de esta Recomendación es voluntaria. Ahora bien, la Recomendación puede contener ciertas disposiciones obligatorias (para asegurar, por ejemplo, la aplicabilidad o la interoperabilidad), por lo que la observancia se consigue con el cumplimiento exacto y puntual de todas las disposiciones obligatorias. La obligatoriedad de un elemento preceptivo o requisito se expresa mediante las frases "tener que, haber de, hay que + infinitivo" o el verbo principal en tiempo futuro simple de mandato, en modo afirmativo o negativo. El hecho de que se utilice esta formulación no entraña que la observancia se imponga a ninguna de las partes.

PROPIEDAD INTELECTUAL

La UIT señala a la atención la posibilidad de que la utilización o aplicación de la presente Recomendación suponga el empleo de un derecho de propiedad intelectual reivindicado. La UIT no adopta ninguna posición en cuanto a la demostración, validez o aplicabilidad de los derechos de propiedad intelectual reivindicados, ya sea por los miembros de la UIT o por terceros ajenos al proceso de elaboración de Recomendaciones.

En la fecha de aprobación de la presente Recomendación, la UIT no ha recibido notificación de propiedad intelectual, protegida por patente, que puede ser necesaria para aplicar esta Recomendación. Sin embargo, debe señalarse a los usuarios que puede que esta información no se encuentre totalmente actualizada al respecto, por lo que se les insta encarecidamente a consultar la base de datos sobre patentes de la TSB.

© UIT 2006

Reservados todos los derechos. Ninguna parte de esta publicación puede reproducirse por ningún procedimiento sin previa autorización escrita por parte de la UIT.

ÍNDICE

	Página
1 Alcance y aplicación.....	1
1.1 Evaluación de la calidad de funcionamiento de la red	1
1.2 Evaluación de la calidad de funcionamiento de la aplicación.....	2
1.3 Modelos de percepción.....	2
1.4 Marco de modelos en el proceso de evaluación de la calidad de funcionamiento de extremo a extremo	2
2 Referencias	3
3 Abreviaturas, siglas o acrónimos	3
4 Evaluación de la calidad de servicio del usuario de extremo	4
Anexo A – Modelo de opinión para las aplicaciones de navegación por la web.....	4
A.1 Alcance	4
A.2 Introducción.....	4
A.3 Experimento subjetivo de calidad de navegación por la web y resultados	5
A.4 Resultados del modelado	9
A.5 Calidad percibida de sesiones de una página de navegación por la web y de eventos de temporización simple.....	12
A.6 Referencias al anexo A.....	14
Apéndice I – Estimaciones de la calidad de funcionamiento de la red con información limitada	14
I.1 Introducción.....	14
I.2 Conexión de referencia.....	14
I.3 Concatenación de los valores de calidad de transferencia de paquetes.....	15
I.4 Anchura de banda de estrechamiento	16
I.5 Tiempo de toma de contacto.....	19
I.6 Ejemplo de aplicación: Transacción HTTP.....	20
I.7 Resumen	22
I.8 Referencias	23

Introducción

La calidad de funcionamiento percibida por el usuario de aplicaciones de datos en redes de paquetes depende de muchos factores; a continuación se indican algunos factores clave:

- 1) La calidad de funcionamiento de extremo a extremo de la red de paquetes (por ejemplo, conectividad, pérdida de paquetes y retardos y capacidad de transferencia de paquetes), teniendo en cuenta el diseño de la red y la carga de tráfico de usuario. La calidad de funcionamiento de la red es a menudo el componente dominante de la calidad de transmisión de extremo a extremo.
- 2) La medida en que la aplicación depende de la red de comunicaciones (por ejemplo, el número de intercambios de paquetes necesarios para completar una transacción, el control de flujo para la transferencia de datos) y de otras entidades de red que proporcionen servicios a la aplicación (por ejemplo, servidores de nombre de dominio).
- 3) La calidad de funcionamiento de los dispositivos de soporte ajenos al ámbito del operador u operadores de red (por ejemplo, equipo de usuario, anfitriones).
- 4) Las tareas del usuario y el grado de interacción del usuario con la aplicación.

Los diseñadores de redes tienen en cuenta estos factores para garantizar la satisfacción del usuario. Un modelo de calidad de funcionamiento de la aplicación de datos debe incluir tantos de estos factores como sea posible. Una vez estimada la calidad de funcionamiento de la aplicación, pueden aplicarse modelos de percepción para interpretar el nivel de calidad de funcionamiento de extremo a extremo.

En esta Recomendación se supone que el lector es capaz de aportar un mínimo de detalles con respecto a cada uno de los factores clave anteriores, y de utilizar un marco de herramientas (muchas de las cuales se especifican aquí) para estimar la calidad de funcionamiento de extremo a extremo.

Recomendación UIT-T G.1030

Estimación de la calidad de funcionamiento de extremo a extremo en redes IP para aplicaciones de datos

1 Alcance y aplicación

Esta Recomendación cubre el proceso de estimación de la calidad de funcionamiento de extremo a extremo de las aplicaciones que funcionan en redes IP, utilizando:

- la calidad de funcionamiento de la red IP considerada, basándose en mediciones pertinentes o en resultados de la modelización de red;
- especificaciones de la aplicación considerada, desde el punto de vista de sus protocolos directores con las opciones especificadas o de un modelo de la aplicación que tenga como entradas la calidad de funcionamiento de la red y de los aparatos del cliente y produzca como resultado un criterio de medición esencial de la calidad de funcionamiento de la aplicación (por ejemplo, el tiempo de descarga de un fichero);
- un modelo de percepción para las aplicaciones consideradas, que permita interpretar la calidad de funcionamiento de la aplicación como una estimación de la calidad experimentada por un grupo típico de usuarios.

La figura 1 ilustra el proceso general para desarrollar una estimación de la calidad de funcionamiento de extremo a extremo.

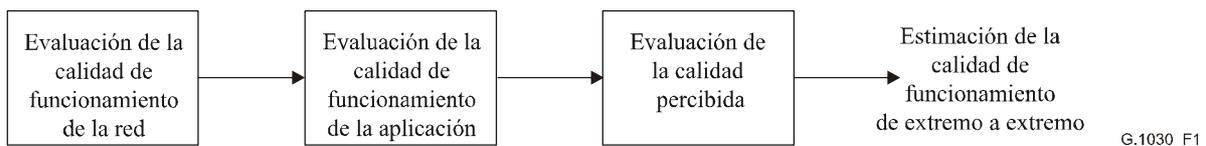


Figura 1/G.1030 – Proceso para obtener la calidad de funcionamiento de extremo a extremo

Se señala que los pasos necesarios para obtener la calidad de funcionamiento de la red y la de la aplicación pueden combinarse en algunos casos, por ejemplo cuando una simulación proporciona un medio para medir la calidad de funcionamiento de una sesión determinada o de un conjunto de sesiones.

1.1 Evaluación de la calidad de funcionamiento de la red

La calidad de funcionamiento de la red puede evaluarse en términos de los parámetros de calidad de funcionamiento de la transferencia de paquetes definidos en la Rec. UIT-T Y.1540, y en otras normas pertinentes (por ejemplo, las RFC elaboradas por el Grupo de Trabajo del IETF sobre métricas de calidad de funcionamiento de IP). Además de las métricas usuales de la calidad de funcionamiento de la transferencia en un solo sentido (unidireccional), la evaluación más directa de la calidad de funcionamiento de la aplicación es posible a veces, con criterios de medición en ambos sentidos (ida y vuelta).

Existen dos fuentes principales de información sobre la calidad de funcionamiento de la red, las mediciones y el modelado.

Las mediciones en la red permiten que el evaluador trate la red como una caja negra, y producen información que puede ser útil en el resto de modelado. No obstante, hay varias consideraciones importantes para el diseño de las mediciones:

- 1) La disciplina en emisión debe en algunos casos concordar con la aplicación considerada. Por ejemplo, el control de flujo del protocolo TCP responde a las condiciones de la red y tiende a llenar la cola en el estrechamiento (bottleneck), incrementando así el retardo por encima de lo que podría medirse en caso contrario.
- 2) La carga sin mediciones debe ser similar, por su tamaño y sus características, a las condiciones en las que han de aplicarse las estimaciones de calidad de funcionamiento. Las mediciones en una red no cargada no son particularmente útiles.

El modelado de la red puede proporcionar la caracterización de la calidad de funcionamiento necesaria cuando la red no esté aún plenamente construida (por ejemplo, los nodos y los enlaces están instalados, pero no está en funcionamiento un protocolo esencial), o cuando no puedan alcanzarse las consideraciones esenciales para las mediciones. Hay muchas herramientas de modelado, incluyendo productos comerciales y herramientas de investigación del dominio público. Las herramientas de modelado exigen bastante pericia e información sobre la red considerada para conseguir los resultados previstos. Como una alternativa a este modelado de red bastante precisa, el apéndice I propone métodos simplificados que pueden utilizarse para estimar la calidad de funcionamiento de extremo a extremo de una red IP. No obstante, la precisión de este método dependerá en gran medida de la precisión de la información disponible.

1.2 Evaluación de la calidad de funcionamiento de la aplicación

Los modelos de aplicación utilizan como entradas las estimaciones de la calidad de funcionamiento de la red y la información en la que se describe la calidad de funcionamiento del dispositivo de aplicación, y producen como resultados una o varias métricas claves de la calidad de funcionamiento de la aplicación.

En la Rec. UIT-T G.1040 se especifica uno de estos modelos de calidad de funcionamiento de aplicación, para intercambios de paquetes conversacionales típicos de las transacciones con tarjeta de crédito y otras transacciones que se realizan en puntos de venta.

Los modelos que se describen en el apéndice I ofrecen resultados precisos para las transferencias de ficheros de larga duración de vida que emplean el servicio fiable de transferencia de bytes del TCP, siempre que las entradas sean correctas.

1.3 Modelos de percepción

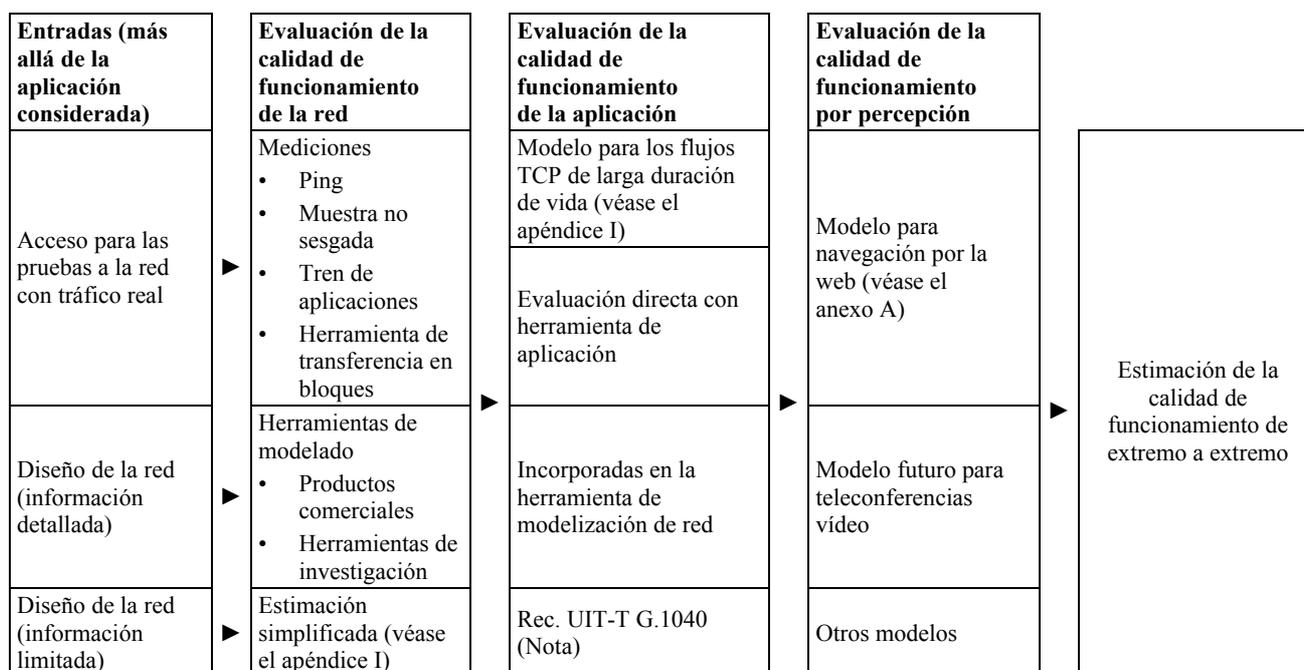
En esta Recomendación se presentan modelos de percepción para obtener la calidad de funcionamiento de extremo a extremo deseada.

En el anexo A se especifica un modelo para la navegación por la web.

Cuando se disponga de modelos adicionales, se incorporarán como anexos a esta Recomendación.

1.4 Marco de modelos en el proceso de evaluación de la calidad de funcionamiento de extremo a extremo

La figura 2 ilustra diversas alternativas del proceso de estimación de la calidad de funcionamiento de extremo a extremo de las aplicaciones en redes IP; en ella, se indica que se dispone de muchas opciones para completar el proceso, aunque en la práctica el evaluador debe combinar opciones que sean coherentes con el objetivo de una estimación de extremo a extremo (y entre sí).



NOTA – La Rec. UIT-T G.1040 proporciona la contribución de la red a la calidad de funcionamiento de las aplicaciones de transacciones. No proporciona la estimación completa de la calidad de funcionamiento de las aplicaciones, pero sí da una visión orientada a la aplicación de la calidad de funcionamiento de la red.

Figura 2/G.1030 – Marco para el proceso de evaluación de la calidad de funcionamiento IP de extremo a extremo

2 Referencias

Las siguientes Recomendaciones del UIT-T y otras referencias contienen disposiciones que, mediante su referencia en este texto, constituyen disposiciones de la presente Recomendación. Al efectuar esta publicación, estaban en vigor las ediciones indicadas. Todas las Recomendaciones y otras referencias son objeto de revisiones por lo que se preconiza que los usuarios de esta Recomendación investiguen la posibilidad de aplicar las ediciones más recientes de las Recomendaciones y otras referencias citadas a continuación. Se publica periódicamente una lista de las Recomendaciones UIT-T actualmente vigentes. En esta Recomendación la referencia a un documento, en tanto que autónomo, no le otorga el rango de una Recomendación.

- Recomendación UIT-T G.1010 (2001), *Categorías de calidad de servicio para los usuarios de extremo de servicios multimedia.*
- Recomendación UIT-T G.1040 (2006), *Contribución de la red al tiempo de transacción.*
- Recomendación UIT-T Y.1540 (2002), *Servicio de comunicación de datos con protocolo Internet – Parámetros de calidad de funcionamiento relativos a la disponibilidad y la transferencia de paquetes del protocolo Internet.*

3 Abreviaturas, siglas o acrónimos

En esta Recomendación se utilizan las siguientes abreviaturas, siglas o acrónimos.

IP Protocolo Internet (*Internet protocol*)

QoS Calidad de servicio (*quality of service*)

TCP Protocolo de control de transmisión (*transmission control protocol*)

4 Evaluación de la calidad de servicio del usuario de extremo

Tras evaluar la calidad física de extremo a extremo, es indispensable que los planificadores de redes evalúen la calidad de servicio del usuario de extremo basándose en aspectos de percepción.

La Rec. UIT-T G.1010 especifica categorías de QoS para aplicaciones de datos, voz y vídeo. Además, en el anexo A se presenta un modelo de opinión que es útil para estimar la calidad, percibida por el usuario, de las aplicaciones de navegación por la web, lo cual hace posible una evaluación basada en valores de retardo de extremo a extremo medidos o planificados teniendo en cuenta la opinión de los clientes.

Anexo A

Modelo de opinión para las aplicaciones de navegación por la web

A.1 Alcance

Este anexo proporciona un modelo para establecer una correspondencia entre los tiempos de respuesta y de descarga (download) medidos en la red o calculados a partir del tiempo de transacción HTTP y la calidad percibida de una sesión de navegación por la web. El modelo se basa en experimentos en los que se manipulan los tiempos de respuesta y descarga de una sesión de navegación por la web [1], [2]. El alcance del modelo se limita actualmente a sesiones de navegación por la web que constan de dos pasos: un primero en el que se efectúa una petición de búsqueda y un segundo en el que se muestra una página de resultados. También se proporciona una extensión simple para eventos de temporización, en la que se modela la influencia de la espera de una página simple.

A.2 Introducción

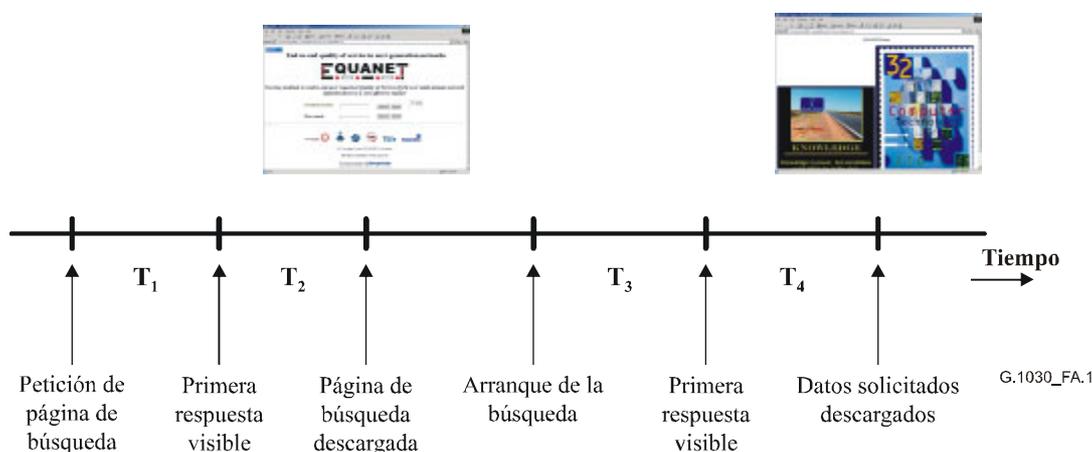
Una observación importante al modelar la calidad percibida durante la navegación por la web es el hecho de que el tiempo máximo de la sesión *esperado* será predominante en la calidad percibida. Si se espera un tiempo de sesión de 100 segundos, la calidad percibida de una sesión de 10 segundos será mucho mayor que si se espera un tiempo de sesión de un segundo. Por consiguiente, el modelo adopta un método que depende del contexto utilizando tres escalas de tiempo diferentes, 6, 15 y 60 segundos, que corresponden a contextos de red rápido, medio y lento, respectivamente.

En general, la percepción de la calidad relativa al tiempo de respuesta puede clasificarse con arreglo a las tres regiones de percepción siguientes [3]:

- 1) **Experiencia instantánea:** 0,1 segundo es aproximadamente el límite para tener la sensación de que el sistema está reaccionando instantáneamente, lo que representa un límite importante para los servicios conversacionales (por ejemplo, el de charla (*chatting*)).
- 2) **Experiencia ininterrumpida:** Un segundo es aproximadamente el límite para que el usuario mantenga ininterrumpido el curso de su pensamiento, aunque pierda la sensación de que el servicio está funcionando directamente, lo que representa un límite importante para los servicios interactivos (por ejemplo, los juegos).
- 3) **Pérdida de atención:** 10 segundos es aproximadamente el límite para mantener la atención del usuario en el diálogo. Con retardos mayores, los usuarios desean realizar otras tareas mientras esperan a que el computador termine, por lo que habría que darles una información de cuándo el computador estima que va a terminar. Esta información durante el retardo es especialmente importante si es probable que el tiempo de respuesta varíe considerablemente, ya que los usuarios no sabrán lo que va a suceder.

En lo que concierne a los tiempos de descarga, los usuarios tienden a adoptar su juicio de calidad con relación al tiempo de descarga esperado [4]. Cuando los usuarios reciben información relativa al tiempo de descarga previsto, estarán dispuestos a aceptar tiempos de descarga mayores.

El modelo de esta Recomendación describe la relación entre los diferentes tiempos de respuesta y descarga durante las sesiones de navegación por la web y la calidad percibida correspondiente para un tiempo de sesión *máximo* dado en una configuración de red y de sistema determinada. El modelo puede aplicarse a una amplia gama de configuraciones de redes y de sistemas, así como a los servicios de navegación por la web para una amplia variedad de usuarios. Los experimentos subjetivos que sirvieron de base al modelo, imitaron lo más posible una experiencia de navegación por la web en la vida real. Para el desarrollo del modelo se utilizaron tres experimentos subjetivos de navegación por la web con escalas de tiempo de unos 6, 15 y 60 segundos, que representan contextos de red rápido, moderado y lento respectivamente. En cada sesión, un usuario recupera en primer lugar una página de búsqueda y a continuación una página que muestra los resultados de la búsqueda. En la figura A.1 se muestra el diagrama de tiempos de una sesión de este tipo. Los dos primeros intervalos de tiempo, T_1 y T_2 , representan los tiempos de respuesta y de descarga no interactivas de la página de búsqueda. Los dos segundos intervalos de tiempo, T_3 y T_4 , representan los tiempos de respuesta y de descarga interactivas de la página de resultados.



T_1 representa el tiempo de respuesta no interactiva; se manipuló utilizando escritura de guiones de Java. T_2 representa el tiempo de descarga no interactiva; se manipuló utilizando un manipulador de red. T_3 y T_4 son los equivalentes de la parte interactiva. La suma $T_1+T_2+T_3+T_4$ representa el tiempo (duración) de la sesión.

Figura A.1/G.1030 – Temporizadores del experimento

Debido a la diferencia conocida entre el comportamiento de los expertos entrenados y los no entrenados, denominados usuarios ingenuos, se efectúa una distinción entre estos grupos al desarrollar el modelo. La separación de estos dos grupos permite desarrollar un modelo que predice la calidad de la navegación por la web de una gran población de usuarios.

A.3 Experimento subjetivo de calidad de navegación por la web y resultados

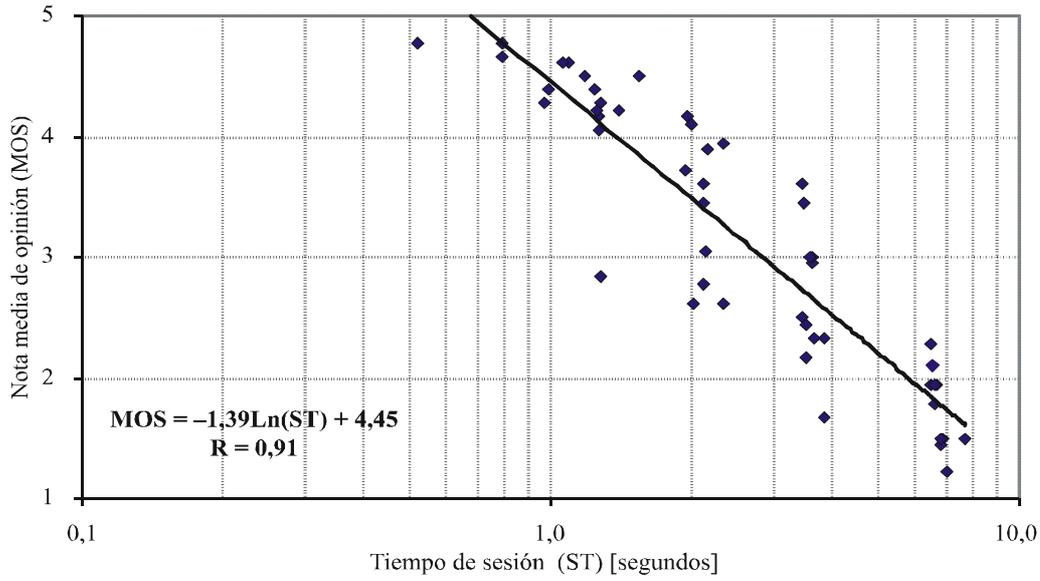
En los experimentos y consecuentemente en el modelo, se utilizó la escala de índices por categorías absolutas UIT-T [5] (escala de cinco puntos con 5: excelente, 4: bueno, 3: regular, 2: mediocre, 1: malo). En cada experimento se presentaron 49 sesiones y cada una constaba de:

- Petición, recuperación y visualización de una página de búsqueda.
- Escritura y envío de un término de búsqueda en esta página.
- Recuperación y visualización de la página de resultados.

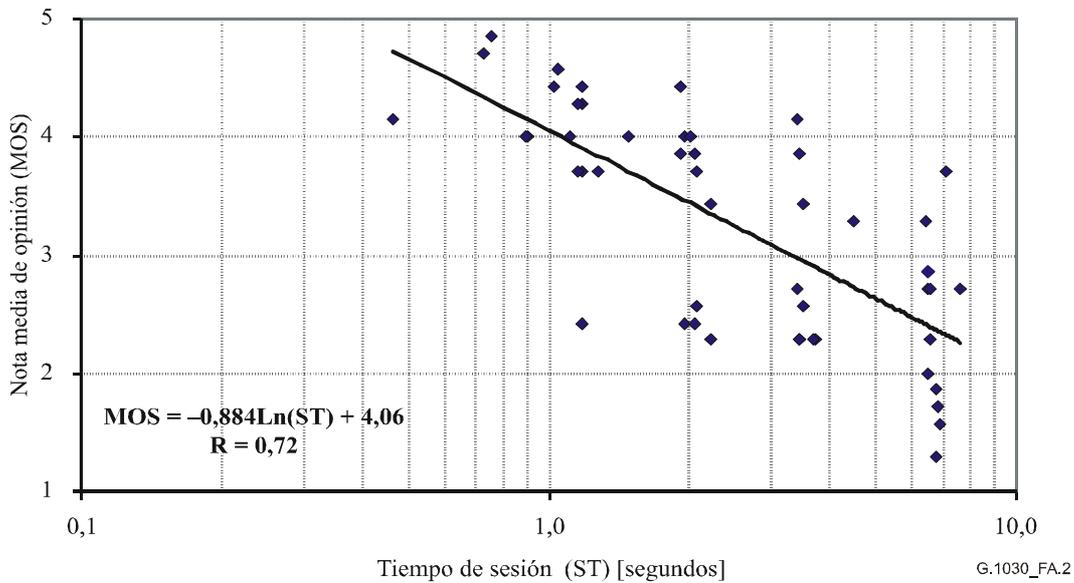
A fin de obtener datos coherentes, en cada sesión se utilizaron exactamente las mismas páginas de resultados y se solicitó a los participantes (usuarios) que escribieran la misma petición de búsqueda en cada sesión. Desde la perspectiva de los participantes, el motor de búsqueda tenía que encontrar, en primer lugar, la página de resultados, la cual se descargaría a continuación. Para cada una de las 49 sesiones, se configuraron diferentes combinaciones de T_1 a T_4 variando la suma $T_1+T_2+T_3+T_4$, es decir, el tiempo de sesión, desde cero a la escala de tiempo para este conjunto de experimentos.

Los resultados globales se presentan en las figuras A.2 a A.4, donde se representa gráficamente el tiempo de sesión (es decir, $T_1+T_2+T_3+T_4$) frente al valor de nota media de opinión (MOS) de los tres experimentos. Para el contexto de larga duración (véase la figura A.4), los resultados de los participantes inexpertos y expertos fueron aproximadamente los mismos y se calcularon valores de MOS para el conjunto de la población. Para los dos contextos de duraciones más cortas (véanse las figuras A.2 y A.3), los participantes inexpertos y expertos se comportaron de manera diferente, y las correlaciones entre los tiempos de sesión y la calidad percibida son significativamente diferentes y en consecuencia se presentan separadamente. Todos los resultados muestran el mismo comportamiento: la calidad percibida disminuye linealmente con el logaritmo del tiempo de sesión. La correlación en el contexto experimental de larga duración es lo bastante alta ($>0,9$) para hacer predicciones fiables de calidad para participantes inexpertos y expertos. En general, las correlaciones por encima de 0,9 están orientadas al modelado sicofísico de la percepción de la calidad [6]. Los resultados muestran también que para el experimento de seis segundos con participantes inexpertos, la correlación entre el tiempo de sesión y la calidad percibida es demasiado baja (0,72) para permitir un modelo simple basado únicamente en el tiempo de sesión.

Resultados con participantes expertos

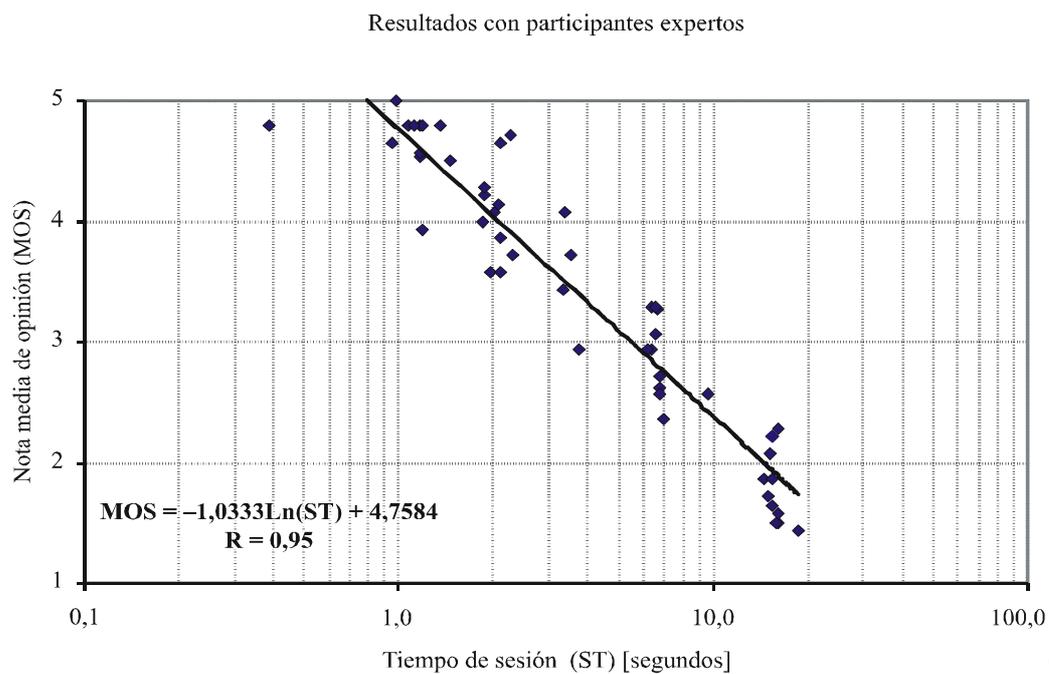
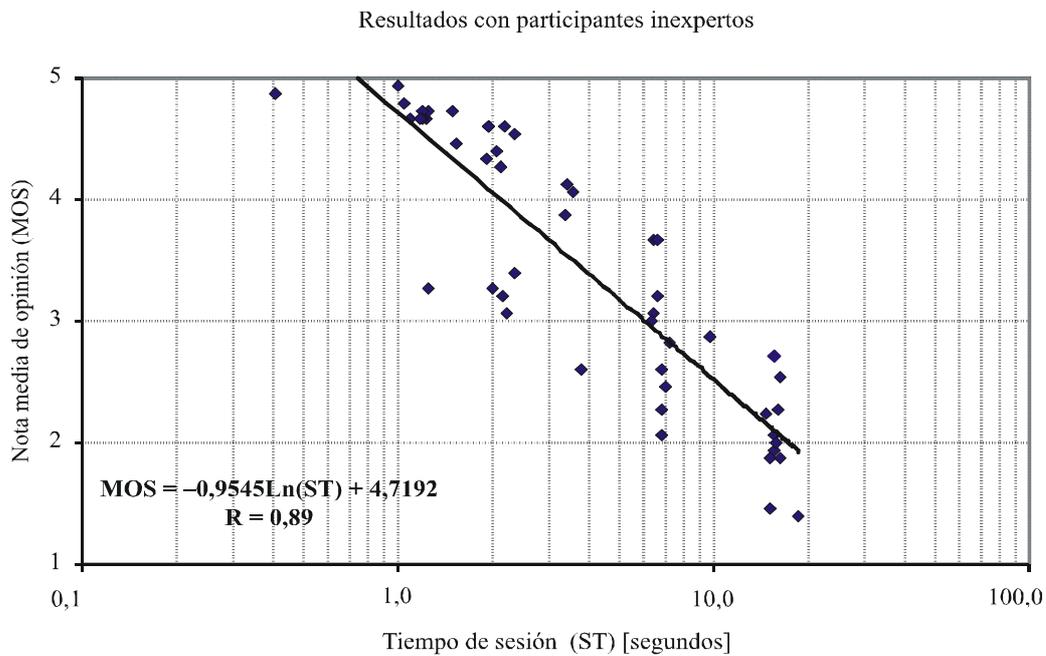


Resultados con participantes inexpertos



Resultados con (siete) participantes inexpertos y (18) expertos y una escala de tiempo de seis segundos. Con participantes inexpertos, la correlación es demasiado baja para permitir predicciones precisas de MOS. Con participantes expertos, la MOS puede predecirse a partir de una interpolación logarítmica del tiempo de sesión entre 0,67 y 12 segundos.

Figura A.2/G.1030 – Resultados de los participantes con una escala de tiempo de seis segundos

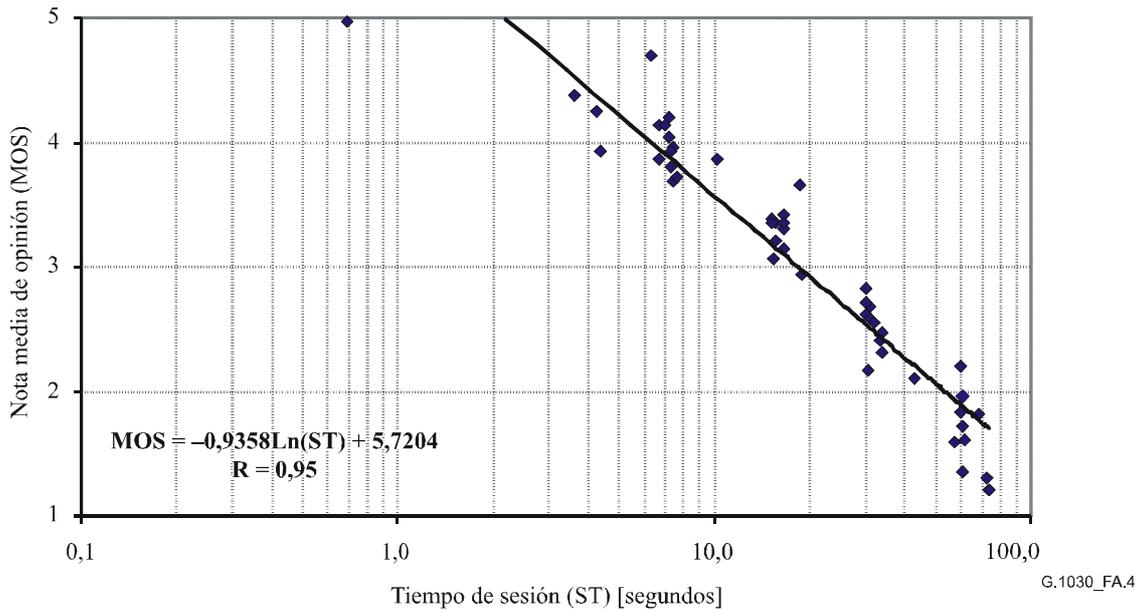


G.1030_FA.3

Resultados con (15) participantes inexpertos y (14) expertos y una escala de tiempo de 15 segundos. Ambos resultados muestran el mismo comportamiento. Con participantes inexpertos, la correlación es ligeramente baja para permitir predicciones precisas de MOS. Con participantes expertos, la MOS puede predecirse a partir de una interpolación logarítmica del tiempo de sesión entre unos 0,79 y 38 segundos.

Figura A.3/G.1030 – Resultados de los participantes con una escala de tiempo de 15 segundos

Resultados con todos los participantes



Resultados con todos los participantes (12 + 17 = 29) y una escala de tiempo de 60 segundos. La MOS puede predecirse a partir de una interpolación logarítmica del tiempo de sesión entre 2,16 y 155 segundos.

Figura A.4/G.1030 – Resultados con todos los participantes con una escala de tiempo de 60 segundos

A.4 Resultados del modelado

A.4.1 Contexto de 60 segundos

En el contexto de 60 segundos, las correlaciones entre el tiempo de sesión y la calidad subjetiva son muy buenas (0,95) tanto para participantes inexpertos como expertos, y la correspondencia entre el tiempo de sesión y la calidad subjetiva puede emplearse directamente en el modelo objetivo. La regresión muestra que con tiempos de sesión de menos de unos dos segundos se obtiene un valor máximo de MOS de cinco (véase la figura A.4), mientras que el valor mínimo de MOS de uno se obtiene con tiempos de sesión mayores que el mayor de los tiempos de sesión en el experimento real. Puede establecerse una correspondencia general entre el tiempo de sesión y la calidad de la navegación por la web en el contexto de larga duración definiendo un tiempo de sesión mínimo (*Mín*) y uno máximo (*Máx*) y utilizando una interpolación logarítmica entre estos tiempos de sesión extremos. Si escribimos $MOS = a - b \cdot \ln(\text{tiempo de sesión})$, utilizamos $MOS = 5$ para $\text{tiempo de sesión} = \text{Mín}$ y $MOS = 1$ para $\text{tiempo de sesión} = \text{Máx}$, para los tiempos de sesión entre *Mín* y *Máx*:

$$MOS = \frac{4}{\ln(\text{Mín} / \text{Máx})} \cdot (\ln(\text{tiempo de sesión}) - \ln(\text{Mín})) + 5 \quad (1)$$

En el experimento de larga duración (véase la figura A.4) la regresión es:

$$MOS = 5,72 - 0,936 \cdot \ln(\text{tiempo de sesión}) \quad (2)$$

acotada entre MOS 1,0 y 5,0.

A.4.2 Contextos de seis y 15 segundos

En los contextos experimentales de seis y 15 segundos, las correlaciones entre el tiempo de sesión y la calidad subjetiva son mucho menores que en el contexto de 60 segundos, y puede elaborarse un

modelo más avanzado para la predicción de la calidad subjetiva con usuarios inexpertos y expertos partiendo de la idea de que con tiempos de sesión de duraciones más cortas el último tiempo de descarga (T_4 en el experimento) tiene una influencia más negativa en la calidad final percibida de navegación por la web que con los demás tiempos de respuesta y descarga (T_1, T_2, T_3 en el experimento). El cuadro A.1 presenta los factores de ponderación que habrán de aplicarse a T_1 a T_4 a fin de obtener una cantidad que tenga la correlación más elevada con los valores de MOS determinados subjetivamente. Esta cantidad, que es el tiempo de sesión ponderado,

$$ST_{ponderado} = WT1 \cdot T_1 + WT2 \cdot T_2 + WT3 \cdot T_3 + WT4 \cdot T_4$$

puede hacerse corresponder con el valor de MOS utilizando la misma interpolación logarítmica entre los tiempos de sesión mínimo y máximo utilizada en (1):

$$MOS = \frac{4}{\ln(Min / Máx)} \cdot (\ln(ST_{ponderado}) - \ln(Min)) + 5 \quad (3)$$

El cuadro A.1 muestra que en el contexto de duración más corta, la influencia del último tiempo de descarga es más del doble que la de los otros tiempos de descarga y respuesta. También muestra un comportamiento significativamente diferente con participantes inexpertos y expertos; la ponderación óptima con participantes inexpertos muestra una influencia mayor del último tiempo de descarga que las ponderaciones óptimas con participantes expertos. Con participantes inexpertos, la influencia del gran tiempo de descarga es más del cuádruple que la de los otros tiempos de descarga y respuesta. Con participantes expertos, es de aproximadamente un factor de dos, mientras que la mejor ponderación general muestra una influencia de alrededor del triple.

Cuadro A.1/G.1030 – Ponderación óptima del modelo para T_1, T_2, T_3 y T_4 con las correlaciones del modelo asociadas entre la temporización objetiva y los resultados subjetivos de MOS

	WT1	WT2	WT3	WT4	<i>Min</i>	<i>Máx</i>	Correlación
6 s con expertos	0,56	0,84	0,80	1,80			0,97
6 s con inexpertos	0,37	0,40	0,60	2,63			0,93
6 s todos juntos	0,47	0,60	0,71	2,22	0,62	13,5	0,95
15 s con expertos	0,63	0,77	1,11	1,49			0,98
15 s con inexpertos	0,48	0,70	0,88	1,95			0,96
15 s todos juntos	0,54	0,72	0,98	1,76	0,81	39	0,97
60 s con expertos	0,84	0,77	1,22	1,18			0,99
60 s con inexpertos	0,64	1,01	1,12	1,24			0,98
60 s todos juntos	0,73	0,90	1,16	1,22	2,22	151	0,98
<i>60 s todos juntos, sin ponderación, véase la figura A.4</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>1,00</i>	<i>2,16</i>	<i>155</i>	<i>0,95</i>
NOTA – La suma de los coeficientes de ponderación está normalizada a 4,0 a fin de poder comparar los tiempos de sesión normales ($T_1+T_2+T_3+T_4$) con los tiempos de sesión ponderados. La ponderación utilizada en el modelo normalizado y los tiempos <i>Min</i> y <i>Máx</i> empleados en la ecuación (3) se muestran en negritas. Con contextos de largos tiempos de sesión, >60 segundos, puede utilizarse el modelo simple sin ponderación (negritas en bastardilla).							

El cuadro A.1 muestra también que en el contexto de duración media, los factores de ponderación con participantes inexpertos y expertos así como los factores de ponderación global se encuentran entre los factores de ponderación para los experimentos de los contextos de duración corta y larga. Esto demuestra la validez del método de ponderación, permitiendo así una interpolación entre los diferentes *tiempos de contexto de sesión experimentales* (es decir, las escalas de tiempo de 6, 15 y 60 segundos) con el objeto de obtener ponderaciones para otros tiempos de contexto.

A.4.3 Resumen

Esta Recomendación normaliza un modelo simple para el contexto de duración de 60 segundos que permite evaluar sesiones de navegación por la web para las cuales el tiempo máximo de sesión es de unos 155 segundos utilizando la ecuación (2). Normaliza además tres modelos avanzados empleando las mejores ponderaciones generales del cuadro A.1 en combinación con las siguientes correspondencias entre el tiempo de sesión ponderado y la calidad de navegación percibida expresada mediante notas medias de opinión:

$$MOS = 4,38 - 1,30 \cdot \ln(\text{Tiempo de sesión ponderado}) \quad (4)$$

acotada entre MOS 1,0 y 5,0 para sesiones de corta duración.

$$MOS = 4,79 - 1,03 \cdot \ln(\text{Tiempo de sesión ponderado}) \quad (5)$$

acotada entre MOS 1,0 y 5,0 para sesiones de duración media.

$$MOS = 5,76 - 0,948 \cdot \ln(\text{Tiempo de sesión ponderado}) \quad (6)$$

acotada entre MOS 1,0 y 5,0 para sesiones de larga duración.

Los resultados de los ajustes de regresión de las ecuaciones (4), (5) y (6) se presentan en las figuras A.5 a A.7.

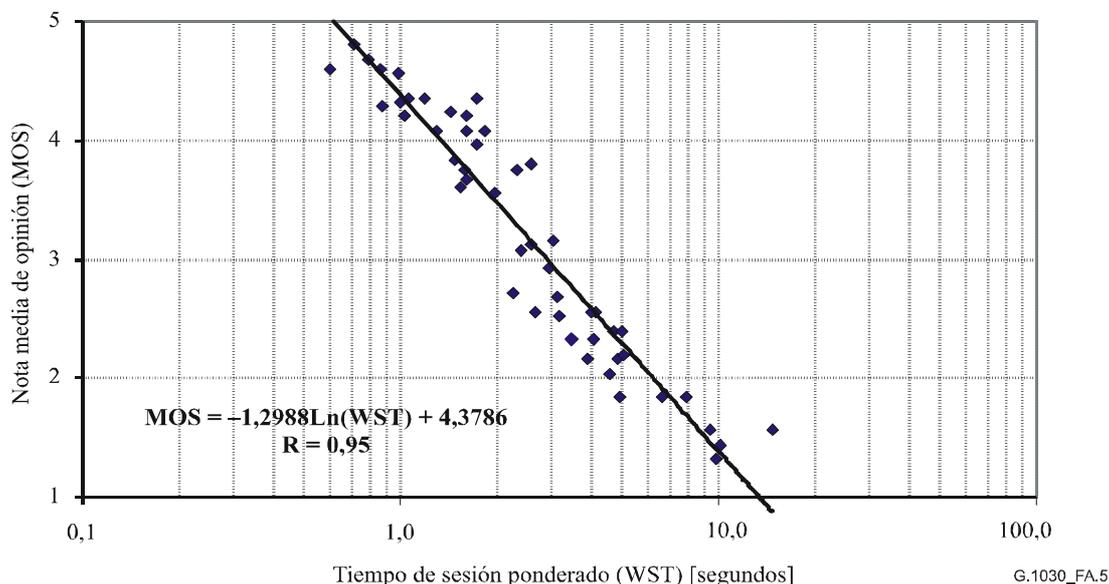


Figura A.5/G.1030 – Modelo y datos para todos los participantes con una escala de tiempo de seis segundos

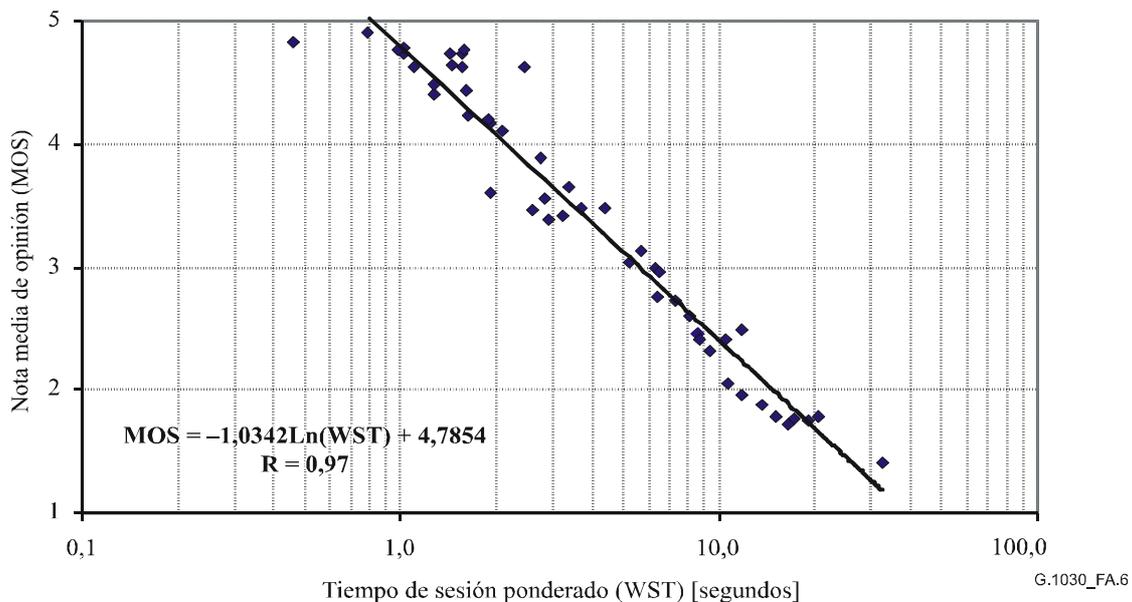


Figura A.6/G.1030 – Modelo y datos para todos los participantes con una escala de tiempo de 15 segundos

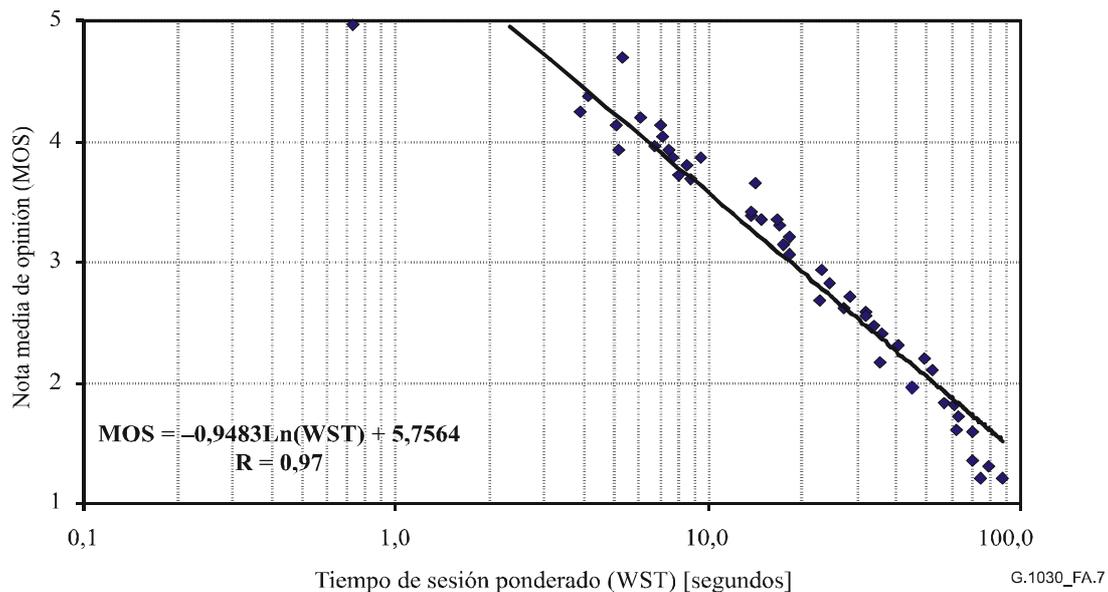


Figura A.7/G.1030 – Modelo y datos para todos los participantes con una escala de tiempo de 60 segundos

A.5 Calidad percibida de sesiones de una página de navegación por la web y de eventos de temporización simple

Basándose en los datos experimentales que representaron el punto de arranque para obtener las ecuaciones (1) y (3), podemos obtener también una relación entre el tiempo de sesión de las sesiones de navegación por la web de una sola página web y la calidad percibida. Estamos también interesados en las sesiones de una sola página en las que el tiempo de descarga siempre es igual a cero. Esta situación se produce cuando la información recuperada aparece al usuario instantáneamente tras un cierto tiempo de espera. Esta última relación hace corresponder un evento

de temporización simple con la calidad percibida subjetivamente expresada mediante una nota MOS.

La obtención de las correspondencias de una página comienza con la observación de que la calidad decrece linealmente con el logaritmo del tiempo de la sesión entre unos tiempos de sesión mínimo y máximo para los que las notas MOS son 5,0 y 1,0 respectivamente (véanse las figuras A.2 a A.7). A partir de los datos de navegación de dos páginas que se presentan en las figuras A.2 a A.4, podemos ver que el tiempo de sesión mínimo en los tres contextos experimentales subjetivos que se utilizaron varía entre aproximadamente 0,7 y 2,2 y aumenta cuando lo hace la duración máxima en el experimento (véanse los resultados en el cuadro A.2).

Cuadro A.2/G.1030 – Tiempos Mín y Máx en los modelos sin ponderación en el experimento de navegación de dos páginas

	<i>Mín (seg)</i>	<i>Máx (seg)</i>
6 s con expertos	0,67	12
15 s con expertos	0,79	38
60 s todos juntos	2,16	155

A partir de estos resultados podemos definir un tiempo de sesión mínimo (*Mín*), para el que se obtiene una nota MOS de 5,0, $Mín = 0,011Máx + 0,47$, donde *Máx* representa el tiempo de sesión máximo previsto. Esto permite una correspondencia general entre el tiempo de sesión y la nota MOS para el experimento de dos páginas con cualquier duración máxima esperada (*Máx*) de la sesión de navegación por la web de dos páginas:

$$MOS_{2-p\acute{a}ginas} = \frac{4}{\ln((0,011Máx + 0,47) / Máx)} \cdot (\ln(Tiempo\ de\ sesi\acute{o}n) - \ln(0,011Máx + 0,47)) + 5 \quad (7)$$

A partir de los datos experimentales que se han presentado podemos estimar que esta relación se mantendrá probablemente para todas las sesiones de navegación por la web de dos páginas entre aproximadamente 10 y 200 segundos.

En las sesiones arbitrarias de una página que sólo tienen un tiempo de respuesta T_1 y un tiempo de descarga T_2 (véase la figura A.1) la calidad más alta se percibe para un valor mínimo de aproximadamente la mitad del valor obtenido en las sesiones de dos páginas. Esto lleva a una definición del tiempo de sesión mínimo $Mín = 0,005Máx + 0,24$ mientras que la correspondencia con los valores MOS viene dada por la ecuación (8):

$$MOS_{1-p\acute{a}gina} = \frac{4}{\ln((0,005Máx + 0,24) / Máx)} \cdot (\ln(Tiempo\ de\ sesi\acute{o}n) - \ln(0,005Máx + 0,24)) + 5 \quad (8)$$

A partir de los datos experimentales que se han presentado podemos estimar que es probable que esta relación se mantenga para todas las sesiones de navegación por la web de una página entre unos cinco y 100 segundos.

Análogamente, en sesiones arbitrarias de una página en las que el tiempo de descarga siempre es igual a 0, o para cualquier evento de temporización simple obtenemos:

$$MOS_{evento\ de\ temporizaci\acute{o}n\ simple} = \frac{4}{\ln((0,003Máx + 0,12) / Máx)} \cdot (\ln(Tiempo\ de\ sesi\acute{o}n) - \ln(0,003Máx + 0,12)) + 5 \quad (9)$$

A partir de los datos experimentales que se han presentado podemos estimar que es probable que esta relación se mantenga para todos los eventos de temporización simple entre unos tres y 50 segundos. Obsérvese que para dicha interacción simple, el tiempo mínimo es de 0,12 segundos, que corresponde al umbral de percepción instantánea [3].

Debe señalarse que los modelos anteriores para la calidad percibida de las sesiones de navegación por la web de una página y de los eventos de temporización simple deberían validarse con experimentos reales.

A.6 Referencias al anexo A

- [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*, noviembre de 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 24 de marzo de 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] Recomendación UIT-T P.800 (1996), *Métodos de determinación subjetiva de la calidad de transmisión*.
- [6] Recomendación UIT-T P.862 (2001), *Evaluación de la calidad vocal por percepción: Un método objetivo para la evaluación de la calidad vocal de extremo a extremo de redes telefónicas de banda estrecha y códecs vocales*.

Apéndice I

Estimaciones de la calidad de funcionamiento de la red con información limitada

I.1 Introducción

Este apéndice proporciona información sobre los métodos simplificados que pueden utilizarse para estimar la calidad de funcionamiento de extremo a extremo en una red IP cuando la información disponible es limitada. La precisión de las estimaciones obtenidas mediante estos métodos depende en gran medida de la calidad de la información de entrada. El material de este apéndice está sujeto a cambios en función de estudios y evaluación ulteriores.

I.2 Conexión de referencia

En esta conexión de referencia es necesario considerar los cortafuegos, los mandatarios (para la mejora de la web y/o la calidad de funcionamiento), los traductores de direcciones de red (NAT) y los conmutadores de equilibrado de la carga, ya que estos dispositivos existen actualmente en muchos hogares, empresas y redes IP gestionadas. Estos "dispositivos intermedios" desempeñan diversas funciones, pero impiden la transparencia de extremo a extremo, que era un aspecto deseable de la arquitectura de red IP. La figura I.1 muestra una conexión de referencia apropiada.

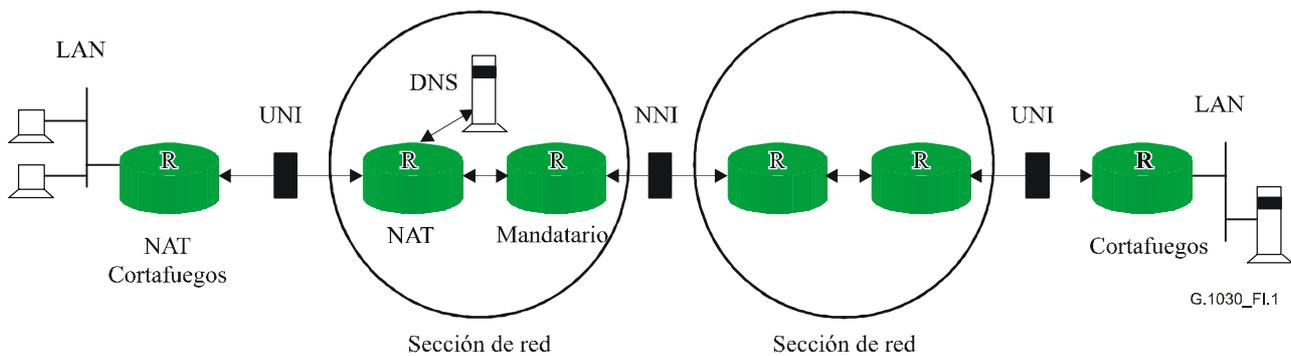


Figura I.1/G.1030 – Conexión de referencia con ejemplos de dispositivos intermedios

Esta extensión permite la evaluación de la calidad de funcionamiento de los protocolos de señalización tales como el SIP utilizando el mismo marco general de otras aplicaciones IP.

I.3 Concatenación de los valores de calidad de transferencia de paquetes

La introducción de dispositivos intermedios significa que habrá más secciones individuales en el trayecto de extremo a extremo, creando la necesidad de fórmulas de acumulación precisas.

I.3.1 Retardo

Los retardos medios de las secciones individuales son aditivos. Sin embargo, puede observarse que el retardo medio representa el centro de gravedad de la distribución de la variación del retardo observado a menudo en el transporte de paquetes. Cuando el retardo de transferencia de paquetes a través de una sección de red (o el tiempo de procesamiento en un anfitrión) se representa por la media, una muestra simple de la distribución del retardo está siendo sustituida por su valor esperado. En esta parte son apropiados los promedios de tiempo del retardo, debido a que cada transacción muestreará la distribución del retardo subyacente en muchos instantes distintos durante el intervalo de transacción.

I.3.2 Pérdida

La característica de tasa de pérdida de extremo a extremo (IPLR, *end-to-end loss ratio*) es esencialmente la calidad de funcionamiento de UNI a UNI. Puede suponerse que las probabilidades de pérdida de los terminales de extremo y de los anfitriones son despreciables.

El método propuesto para la concatenación de las IPLR es invertir la probabilidad de la transferencia de paquetes satisfactoria a través de n secciones de red, de la siguiente manera:

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

Esta ecuación se basa en la teoría de probabilidades condicionales y supone que las probabilidades de pérdida en cada sección de red son independientes. Para un trayecto de UNI a UNI con dos redes, A y B, con probabilidades de pérdida p_A y p_B :

$$\begin{aligned} \text{Prob}\{\text{éxito en ambas redes}\} &= \text{Prob}\{\text{éxito en B} \mid \text{éxito en A}\} \times \text{Prob}\{\text{éxito en A}\} \\ &= (1 - p_B) \times (1 - p_A) \end{aligned}$$

Las probabilidades de pérdida que se obtienen de manera empírica para las redes individuales (p_A y p_B) son equivalentes a las probabilidades condicionales de éxito en cualquiera de las redes precedentes. Éstas necesitan un evento de entrada de paquetes para comenzar la operación de medición.

I.3.2.1 Aditividad de la pérdida de paquetes en modo ráfaga

Queda en estudio.

I.3.3 Variación del retardo

Las aplicaciones IP tradicionales y muchas otras aplicaciones de datos son menos sensibles a la variación del retardo que las aplicaciones que exigen velocidades continuas de transferencia de paquetes (las isócronas o las aplicaciones en tiempo real tales como VoIP). Por este motivo, las aplicaciones IP tradicionales se caracterizan a veces por sus trenes de paquetes "elásticos", que pueden comprimirse o alargarse durante la transferencia sin mayores efectos. Cuando el espaciado de los paquetes de VoIP cambia durante el transporte, la aplicación exige memorias intermedias adicionales para poder restablecer la capacidad de transferencia continua de paquetes, y la memoria intermedia por sí misma representa una fuente de retardo con estos trenes de paquetes "no elásticos".

Es difícil concatenar la variación del retardo de varias secciones y procesadores de red de una manera precisa, ya que raramente se conoce la distribución completa del retardo. Sin embargo, en la cláusula 8/Y.1541 se presenta un método de aproximación razonable.

I.4 Anchura de banda de estrechamiento

La utilización de la anchura de banda de acceso como el factor limitador cuando se calcula el tiempo de transferencia de datos de extremo a extremo puede dar lugar a una visión demasiado optimista del caudal de tráfico.

I.4.1 Consideraciones clave

Una aproximación más precisa del tiempo de transferencia de datos real de un fichero o mensaje de tamaño fijo puede realizarse incluyendo las siguientes consideraciones en el modelo:

- 1) Debe indicarse el sentido de transferencia. Las velocidades de acceso son asimétricas en muchas de las nuevas tecnologías, tales como las de DSL y de módems de cable, y la diferencia puede ser de un orden de magnitud o más.
- 2) Debe utilizarse la anchura de banda real de estrechamiento. La velocidad de acceso no es necesariamente la del estrechamiento. Uno de los anfitriones puede reducir su velocidad de transmisión (al no poder pasar suficientes datos al proceso TCP, o con límites de ventana), los parámetros de TCP del anfitrión pueden limitar el caudal y dispositivos intermedios tales como los apoderados de mejora de la calidad de funcionamiento pueden modificar los parámetros de TCP para mejorar el caudal, etc.
- 3) La transacción puede encontrar una contienda por recursos compartidos. Debe utilizarse anchura de banda eficaz de estrechamiento implícita en los objetivos de diseño para soportar un determinado número de usuarios simultáneos en un enlace u otros recursos compartidos.
- 4) La anchura de banda de estrechamiento puede variar en función del tiempo. Los límites de anchura de banda producirán una gama de tiempos de transferencia de datos.
- 5) La tara debe calcularse en el punto de estrechamiento. Cada una de las capas por debajo de IP añade alguna tara de encabezamiento. Supóngase un tamaño de cabida útil y calcúlese el porcentaje de tara; esto aumenta aún más el tiempo de transferencia de datos.
- 6) El protocolo TCP no puede lograr una capacidad de transferencia igual a la anchura de banda de estrechamiento. La ecuación de Mathis para TCP da una buena aproximación de la capacidad de transferencia en régimen permanente, pero los factores antes mencionados pueden tener una importancia primordial.

La mayoría de estas cuestiones requieren mayor preparación por parte del usuario del modelo o del planificador de la red. En I.4.2 se presentan los detalles de la aproximación de capacidad de TCP.

I.4.2 Capacidad de TCP

La gran mayoría de las aplicaciones IP tradicionales utilizan los servicios de transferencia de tren de bytes fiables de TCP (protocolo de control de transmisión). Un modelo de comportamiento del control de flujo TCP [TCPCon] viene dado por:

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

donde:

BW representa los datos transferidos por unidad de tiempo (tiempo de ciclo)

MSS es el tamaño del segmento máximo de TCP

RTT es el tiempo medio de ida y vuelta

C es una constante que tiene en cuenta los efectos de la pérdida aleatoria/periódica y de la estrategia de ACK

p es la probabilidad de pérdida de paquetes

Esta ecuación debería considerarse como un límite superior del flujo de información, incluso para las recientes mejoras del TCP, tales como SACK y TCP Reno. La ecuación supone que la conexión TCP tiene un tiempo de vida suficientemente largo como para alcanzar el equilibrio en el estado evitar congestión. Cuando el total de bytes en la transferencia de datos es pequeño, es probable que el control de flujo de TCP no alcance el equilibrio, y el cálculo se vea dominado por el comportamiento de arranque lento. En [Padhye], puede encontrarse una relación más flexible para la capacidad TCP Reno, que incluye los efectos limitadores del tamaño de ventana máximo.

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

donde:

$B(p)$ Modelo aproximado del caudal de TCP [paquetes/s]

W_{\max} Tamaño máximo de memoria intermedia de la ventana del receptor [paquetes]

RTT Tiempo de ida y vuelta [seg]

b Número de paquetes de los que se ha acusado recibo por un ACK recibido

p Probabilidad de que se pierda un paquete

T_0 Temporización para la retransmisión de un paquete del que no se ha acusado recibo (perdido) [seg]

El control de flujo del TCP trata de aumentar su velocidad en emisión (tamaño de la ventana) hasta que encuentra congestión (o un enlace con estrechamiento), y así lo deduce de la pérdida de paquetes. Por consiguiente, parte de la pérdida de paquetes es consecuencia de sondeo de estrechamientos. El control de flujo para evitar la congestión que consiste en reducir a la mitad la ventana cuando se produce una pérdida y a continuación aumentar el tamaño de la ventana en un paquete cada vez hasta que se produce otra pérdida, limita efectivamente el caudal aproximadamente al 75% del valor de cresta del tamaño de la ventana o de la velocidad de transmisión. Se puede modificar esta ecuación para encontrar la pérdida de paquetes debida sólo al sondeo de estrechamientos del TCP.

$$\sqrt{p_{\text{Sondeo}}} \approx \frac{MSS}{RTT} \frac{C}{0,75 \times BW(\text{enlace_con_estrech.})}$$

El cuadro I.1 enumera la pérdida de paquetes inherente asociada a una velocidad particular de enlace con estrechamiento.

Cuadro I.1/G.1030 – Pérdida de paquetes de TCP debida al sondeo de estrechamientos

C	MSS	RTT	BW (enlace)	75% BW	Retardo*BW	Ventana	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	5 120	0,4266667	7,32E+00
0,866	2 048	0,08	128 000	96 000	10 240	5	5,33E-02
0,866	2 048	0,08	64 000	48 000	5 120	2,5	2,13E-01

Cabe observar varios puntos del cuadro I.1:

- La pérdida de paquetes inherente es muy baja cuando el BW del enlace con estrechamiento es de 10 Mbit/s (Ethernet o tecnologías con velocidades mayores). Si la pérdida de paquetes de la red es de 10^{-4} o peor, tenderá a reducir el caudal de conformidad con [TCPCon].
- Para los parámetros dados (RTT de 80 ms, MSS de 12 000 bits (1500 bytes), etc.) aproximadamente uno de cada 100 paquetes se pierden en un enlace T_1 cuando un flujo TCP simple sondea el estrechamiento. Una pérdida de paquetes de la red de 10^{-3} o mejor no tendrá casi consecuencias para el caudal resultante.
- Se muestran el producto retardo*BW y el tamaño de ventana óptimo (producto/MSS) para el trayecto, y se indica que hay dos BW de enlace (128 kbit/s y 64 kbit/s) en los que la ventana es menor que uno y la tasa de pérdida calculada no tiene sentido (>1). En estos casos, está justificado el ajuste por el protocolo del TCP, de manera que se puede reducir el tamaño máximo de segmento, MSS, a 2048 bits y producir un resultado más razonable.
- Se señala que la reducción de BW debido a la tara de capas inferiores no se aborda en el cuadro.

La tasa de pérdida de la red estimada se combinará con la tasa de pérdida de sondeo como se muestra a continuación.

$$p_{Total} = 1 - \{(1 - p_{Red}) \times (1 - p_{Sondeo})\}$$

Cuando la tasa de pérdida de paquetes de la red y de sondeo son de mismo orden de magnitud, es hasta cierto punto pesimista combinarlas y calcular el BW, debido a que algunas de las pérdidas de sondeo no se producirán.

Por último, puede observarse que no se ha tratado en detalle el ajuste de parámetros TCP, excepto para decir que los parámetros pueden limitar el caudal en algunos casos. Si el tamaño máximo de ventana es insuficiente para rellenar el trayecto de ida y vuelta, el caudal se limitará a: ventana \times MSS/RTT [Padhye]. Muchas conexiones TCP nunca experimentan una pérdida de paquetes durante su vida útil, debido a ventanas pequeñas y a la tendencia del protocolo TCP a rellenar la cola frente al estrechamiento (y a incrementar el RTT de algunos paquetes – ésta es la razón por la que se utiliza un RTT promedio).

I.5 Tiempo de toma de contacto

El intercambio de los paquetes puede caracterizarse por dos fases:

- 1) Toma de contacto;
- 2) Transferencia de datos (algunas aplicaciones no tienen esta fase, por ejemplo los terminales de punto de venta para tarjetas de crédito y la señalización de VoIP. Se incluye el intercambio final de paquetes para cerrar la conexión en esta fase, como una simplificación. La matriz 3×3 tradicional trata la desconexión como una fase separada).

Si un paquete o su respuesta se pierden durante la toma de contacto inicial, por lo general, el emisor queda en espera un cierto tiempo específico antes de retransmitir el mensaje. El tiempo de espera o temporización de la retransmisión puede ser del orden de uno a tres segundos, e incrementa significativamente el tiempo de toma de contacto inicial cuando se produce una pérdida.

Si todos los tiempos relacionados con la red (incluidos los causados por las pérdidas de paquetes) se separan de los tiempos de procesamiento del anfitrión, hay dos métricas que corresponden a distintas responsabilidades administrativas (red y CPE/anfitriones).

Esta descomposición es bastante directa para la fase de toma de contacto inicial. Los tiempos de la transferencia de paquetes y las temporizaciones debidas a las pérdidas se atribuyen a las redes.

En el caso del tiempo de transferencia de datos, bastará con dar el tiempo e indicar si el factor limitador fue la pérdida de paquetes de la red, el retardo de la red o el procesamiento/las configuraciones del anfitrión.

I.5.1 Efecto de la pérdida (de paquetes) durante la toma de contacto

La contribución de la red IP al tiempo de toma de contacto inicial puede determinarse como se ilustra a continuación. En primer lugar, consideremos el caso en el que deben intercambiarse ocho mensajes y ocho respuestas, y todas las transferencias de paquetes son fructuosas. La probabilidad de este caso se muestra a continuación.

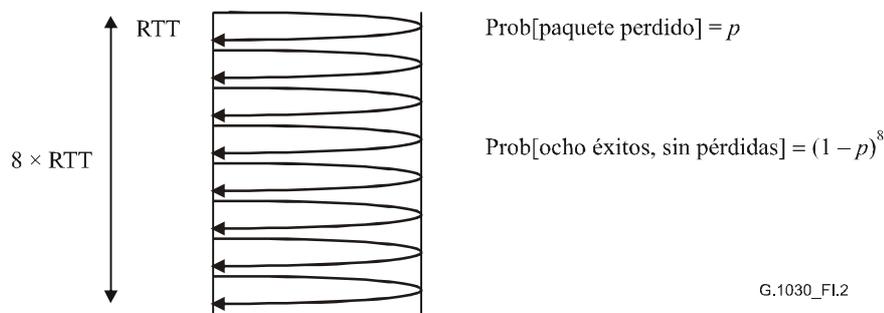


Figura I.2/G.1030 – Toma de contacto con ocho rondas con ida y vuelta (RT) y sin pérdida de paquetes

Éste es un caso simple, pero representa la contribución de la red IP a casi todas las experiencias del cliente cuando la probabilidad de pérdida de paquetes de ida y vuelta, p , es suficientemente baja. Se observa que la probabilidad de pérdida de ida y vuelta es:

$$p = p_{RT} = 1 - \{(1 - p_{un-sentido}) \times (1 - p_{otro-sentido})\}$$

El tiempo necesario para completar los ocho intercambios es de ocho veces el tiempo de ida y vuelta promedio, RTT, más el tiempo de procesamiento en el anfitrión distante (HPT) este total es el tiempo de toma de contacto. Estos componentes pueden calcularse por separado, como:

$$\text{Tiempo_de_toma_de_contacto} = NCTT + \text{Total_HPT}$$

donde NCTT es la contribución de la red al tiempo de transacción, definido en la Rec. UIT-T G.1040.

Obsérvese que el ejemplo de toma de contacto POP3 tiene ocho intercambios de petición/respuesta. Si suponemos que $p = 10^{-3}$, $RTT+HPT = 0,080$ segundos, la temporización de retransmisión ($RTO = 1$ segundo y que hay 350 000 intentos de transacción, tenemos las siguientes probabilidades para cada uno de los casos de pérdida como se muestra en el cuadro I.2.

Cuadro I.2/G.1030 – Tiempo de toma de contacto para los casos con cero a tres paquetes perdidos

Pérdidas	Tiempo de toma de contacto, s	Probabilidad de que se produzcan	Transacciones
0	0,64	0,99202794	347 210
1	1,64	0,00793622	2 778
2	2,64	3,5713E-05	12
3	3,64	1,1904E-07	0

Este ejemplo muestra que la red IP contribuye con menos de un segundo al promedio, el percentil 95 y el percentil 99 del tiempo de toma de contacto para una transacción POP3 con ocho intercambios con ida y vuelta.

I.5.2 Efecto de la pérdida de paquetes durante la transferencia de datos

La pérdida de paquetes exigirá retransmisiones, y con TCP el control de flujo para evitar congestiones reduce la velocidad de transmisión, como se describe en I.4.

I.6 Ejemplo de aplicación: Transacción HTTP

La figura I.3 muestra una transacción HTTP que identifica el tiempo de toma de contacto. El intercambio de paquetes restante constituye el tiempo de transferencia de datos desde la perspectiva del cliente.

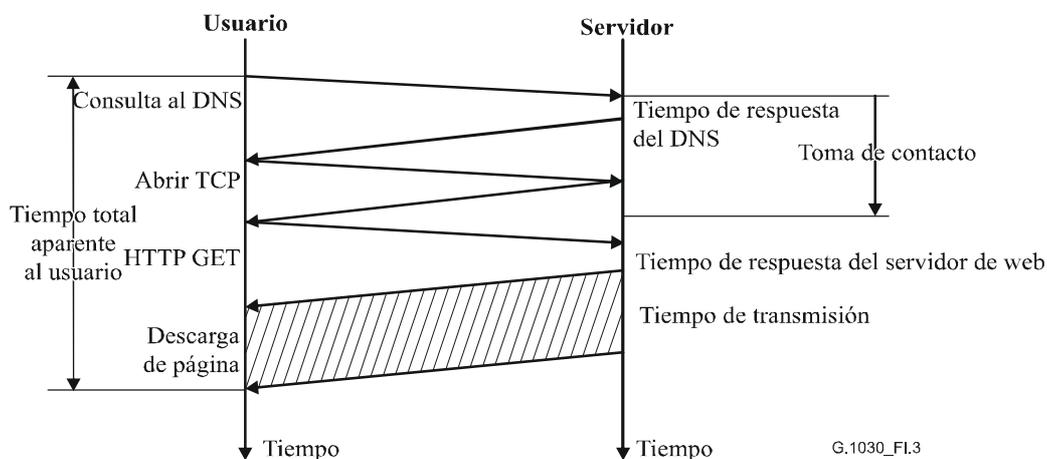


Figura I.3/G.1030 – Intercambio mediante el protocolo HTTP

Se utiliza la conexión de referencia de la figura I.1 en la que se muestran dos secciones de red con traductores NAT, cortafuegos y un apoderado como ejemplo de dispositivos intermedios. Obsérvese que el apoderado no tiene los objetos solicitados almacenados en memoria caché, y que efectúa una petición al servidor de web distante.

I.6.1 Tiempo de toma de contacto

Hay dos peticiones-respuestas efectuadas por diferentes trayectos, la consulta al DNS y la toma de contacto TCP en tres sentidos (donde se temporizan SYN y SYN-ACK; se supone que el último ACK continúa con el HTTP GET y no añade una cantidad de tiempo significativa). Aunque el apoderado divide la conexión TCP, en este ejemplo esto se trata como un retardo adicional en una conexión simple.

Como la mayor parte de la acumulación se considera mediante simples adiciones o multiplicaciones, simplemente tabulamos los valores y mostramos los totales (en lugar de escribir ecuaciones con muchos parámetros nuevos).

En primer lugar, en el cuadro I.3 se tabulan los componentes de la calidad de funcionamiento del DNS.

Cuadro I.3/G.1030 – Consulta al DNS

Cliente al DNS	RTT (red) [ms]	HPT (tiempo de procesamiento) [ms]	RT (1-p)	Por cada BW de flujo
LAN	2		0,99999	5,0E+6
NAT/cortafuegos		4	0,99999	
Enlace 1	10		0,9999	5,0E+6
NAT/encaminador de borde			0,99999	
NAT a enlace de DNS	4		0,99999	10,0E+6
DNS		6	0,99999	
Totales (pérdida invertida)	16	10	0,00015	

Por último, a continuación se tabulan los componentes del establecimiento de la conexión TCP mostrados en el cuadro I.4.

Cuadro I.4/G.1030 – Trayecto cliente-servidor para el tiempo de conexión TCP y el tiempo de transferencia de datos

Cliente a servidor	RTT (red) [ms]	HPT (tiempo de procesamiento) [ms]	RT (1-p)	Por cada BW de flujo
LAN	2		0,99999	5,0E+6
NAT/cortafuegos		4	0,99999	
Enlace 1 de UNI	10		0,9999	5,0E+6
NAT/encaminador de borde			0,99999	
Enlace NAT a mandatario	4		0,99999	10,0E+6
Mandatario		4	0,99999	
Enlace de NNI	20		0,99999	10,0E+6
Sección 2 de la red	14		0,99999	10,0E+6
Enlace 2 de UNI	10		0,99999	3,0E+6

Cliente a servidor	RTT (red) [ms]	HPT (tiempo de procesamiento) [ms]	RT (1-p)	Por cada BW de flujo
Cortafuegos		2	0,99999	
LAN	2		0,99999	5,0E+6
Servidor (Procesamiento de TCP)		1	0,99999	
Totales (pérdida invertida)	62	11	0,00021	

Con el tiempo de consulta al DNS = 16 + 10 = 26 ms, y el tiempo de apertura del TCP de 62 + 11 = 73 ms, el tiempo de toma de contacto es de 99 ms, lo que parece ser válido para más del 99,9% de las transacciones HTTP.

I.6.2 Tiempo de transferencia de datos

Los primeros pasos consisten en determinar si la pérdida entre el cliente y el servidor tiene un papel importante en este cálculo.

Haciendo referencia nuevamente al cuadro I.4, observamos que el BW por flujo más bajo ocurre en el enlace 2 de UNI, a 3 Mbit/s. El producto retardo × BW es 0,073 s × 3 Mbit/s = 219 kbit/s o un tamaño de ventana óptimo de 18,25 paquetes a un tamaño máximo de segmento, MSS, de 12 000 bits (1500 bytes). El protocolo TCP puede ajustarse para dar cabida a ese trayecto, de manera que no existe ninguna restricción asociada al tamaño de la ventana.

Utilizando el BW del enlace estrechamiento, se calcula la tasa de pérdida de paquetes inherente al TCP debida al sondeo mediante:

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

Ya que la pérdida de la red calculada (0,00021) es aproximadamente igual a $p_{\text{Sondeo}}/20$, se concluye que es despreciable, y el caudal del tren de bytes de TCP será $0,75 \times 3 \text{ Mbit/s} = 2,25 \text{ Mbit/s}$.

Si por otro lado, la tasa de pérdida de la red estimada era del mismo orden que la tasa de pérdida de paquetes de sondeo, sería algo pesimista combinarlas y recalculer el BW:

$$p_{\text{Total}} = 1 - \{(1 - p_{\text{Red}}) \times (1 - p_{\text{Sondeo}})\}$$

Suponiendo una página web con gráficos complejos, el tiempo de transferencia de 1 Mbit es de 0,444 segundos. Puede añadirse el tiempo necesario para que el servidor procese el HTTP GET (0,01 segundos) lo que da un total de 0,454 segundos.

I.6.3 Tiempo total aparente al usuario

El total es la suma del tiempo de toma de contacto y del tiempo de transferencia de datos, $0,099 + 0,454 = 0,553$ segundos.

I.7 Resumen

En este apéndice se describe una metodología para estimar la calidad de funcionamiento de extremo a extremo en redes IP. Los resultados obtenidos pueden utilizarse comparándolos con los objetivos de calidad de funcionamiento centrados en el usuario, por ejemplo los de la Rec. UIT-T G.1010, para estimar la satisfacción general del usuario de extremo con relación a las diferentes aplicaciones multimedia.

I.8 Referencias

- Recomendación UIT-T G.1010 (2001), *Categorías de calidad de servicio para los usuarios de extremo de servicios multimedia*.
 - Recomendación UIT-T Y.1541 (2006), *Objetivos de calidad de funcionamiento de red para servicios basados en el protocolo Internet*.
- [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, julio de 1997, ISSN# 0146-4833.
http://www.psc.edu/networking/papers/model_ccr97.ps

SERIES DE RECOMENDACIONES DEL UIT-T

Serie A	Organización del trabajo del UIT-T
Serie D	Principios generales de tarificación
Serie E	Explotación general de la red, servicio telefónico, explotación del servicio y factores humanos
Serie F	Servicios de telecomunicación no telefónicos
Serie G	Sistemas y medios de transmisión, sistemas y redes digitales
Serie H	Sistemas audiovisuales y multimedios
Serie I	Red digital de servicios integrados
Serie J	Redes de cable y transmisión de programas radiofónicos y televisivos, y de otras señales multimedios
Serie K	Protección contra las interferencias
Serie L	Construcción, instalación y protección de los cables y otros elementos de planta exterior
Serie M	Gestión de las telecomunicaciones, incluida la RGT y el mantenimiento de redes
Serie N	Mantenimiento: circuitos internacionales para transmisiones radiofónicas y de televisión
Serie O	Especificaciones de los aparatos de medida
Serie P	Calidad de transmisión telefónica, instalaciones telefónicas y redes locales
Serie Q	Conmutación y señalización
Serie R	Transmisión telegráfica
Serie S	Equipos terminales para servicios de telegrafía
Serie T	Terminales para servicios de telemática
Serie U	Conmutación telegráfica
Serie V	Comunicación de datos por la red telefónica
Serie X	Redes de datos, comunicaciones de sistemas abiertos y seguridad
Serie Y	Infraestructura mundial de la información, aspectos del protocolo Internet y Redes de la próxima generación
Serie Z	Lenguajes y aspectos generales de soporte lógico para sistemas de telecomunicación