

国 际 电 信 联 盟

ITU-T

国际电信联盟
电信标准化部门

G.1030

(11/2005)

G系列：传输系统和媒质、数字系统和网络
服务质量和性能 — 一般和与用户相关的概况

为数据应用评估IP网络端对端性能

ITU-T G.1030建议书

ITU-T G系列建议书

传输系统和媒质、数字系统和网络

国际电话连接和电路	G.100-G.199
所有模拟载波传输系统共有的一般特性	G.200-G.299
金属线路上国际载波电话系统的各项特性	G.300-G.399
在无线接力或卫星链路上传输并与金属线路互连的国际载波电话系统的一般特性	G.400-G.449
无线电话和有线电话的配合	G.450-G.499
传输媒质的特性	G.600-G.699
数字终端设备	G.700-G.799
数字网	G.800-G.899
数字段和数字线路系统	G.900-G.999
服务质量和性能 — 一般和与用户相关的概况	G.1000-G.1999
传输媒质的特性	G.6000-G.6999
经传送网的数据 — 一般概况	G.7000-G.7999
经传送网的以太网概况	G.8000-G.8999
接入网	G.9000-G.9999

欲了解更详细信息，请查阅ITU-T建议书目录。

为数据应用评估IP网络端对端性能

摘 要

本建议书为获得IP网络性能、评估用户应用性能、以及在衡量用户对端对端性能满意程度中应用感知模型提供了工具框架。

分组网中数据应用的用户感知性能取决于很多因素，包括分组网的端对端性能、该应用对通信网的依赖性、终端及其他处于网络运营商范围之外设备的性能、以及用户任务和用户与该应用互动的程度。网络设计者考虑这些因素来确保用户满意。一旦已经评估了应用的性能，则能够用感知模型来解释获得的端对端性能水平。

本建议书假设读者至少能够在一定程度上提供关于以上每个关键因素的详细情况，然后能够能用工具框架来评估端对端的性能。

来 源

ITU-T第12研究组（2005-2008）按照ITU-T A.8建议书规定的程序，于2005年11月29日批准了ITU-T G.1030建议书。

前 言

国际电信联盟（ITU）是从事电信领域工作的联合国专门机构。ITU-T（国际电信联盟电信标准化部门）是国际电信联盟的常设机构，负责研究技术、操作和资费问题，并且为在世界范围内实现电信标准化，发表有关上述研究项目的建议书。

每四年一届的世界电信标准化全会（WTSA）确定 ITU-T 各研究组的研究课题，再由各研究组制定有关这些课题的建议书。

WTSA 第 1 号决议规定了批准建议书须遵循的程序。

属 ITU-T 研究范围的某些信息技术领域的必要标准，是与国际标准化组织（ISO）和国际电工技术委员会（IEC）合作制定的。

注

本建议书为简明扼要起见而使用的“主管部门”一词，既指电信主管部门，又指经认可的运营机构。

遵守本建议书的规定是以自愿为基础的，但建议书可能包含某些强制性条款（以确保例如互操作性或适用性等），只有满足所有强制性条款的规定，才能达到遵守建议书的目的。“应该”或“必须”等其它一些强制性用语及其否定形式被用于表达特定要求。使用此类用语不表示要求任何一方遵守本建议书。

知识产权

国际电联提请注意：本建议书的应用或实施可能涉及使用已申报的知识产权。国际电联对无论是其成员还是建议书制定程序之外的其它机构提出的有关已申报的知识产权的证据、有效性或适用性不表示意见。

至本建议书批准之日止，国际电联尚未收到实施本建议书可能需要的受专利保护的知识产权的通知。但需要提醒实施者注意的是，这可能并非最新信息，因此特大力提倡他们通过下列网址查询电信标准化局（TSB）的专利数据库：<http://www.itu.int/ITU-T/ipr/>。

© 国际电联2006

版权所有。未经国际电联事先书面许可，不得以任何手段复制本出版物的任何部分。

目 录

页码

1	范围与应用	1
1.1	网络性能的评价	1
1.2	应用性能的评价	2
1.3	感知模型	2
1.4	端对端性能评估程序中模型的框架	2
2	参考文献	3
3	缩写词	3
4	最终用户QoS的评定	3
附件 A – Web浏览应用的评价模型		4
A.1	范围	4
A.2	绪论	4
A.3	主观Web浏览质量实验与结果	5
A.4	模拟结果	8
A.5	一页Web浏览进程和单个定时事件的感知质量	11
A.6	附件A的参考文献	12
附录一 – 采用有限信息的网络性能评估		13
I.1	引言	13
I.2	参考连接	13
I.3	包传送性能数值的级联	13
I.4	瓶颈带宽	14
I.5	握手时间	16
I.6	应用举例：HTTP事务	18
I.7	小结	20
I.8	附录一的参考文献	20

引 言

分组网中数据应用的用户感知性能取决于很多因素，一些关键因素列举如下：

- 1) 分组网的端对端性能（例如连通性，丢包和时延，以及包的传送能力），同时顾及网络设计和用户业务量负荷。网络性能常常是端对端传输性能的主导成分。
- 2) 应用对通信网络（例如完成一个事务所需要的包交换次数、数据传送的流控）和其他为该应用提供服务的网络实体（例如域名服务器）的依赖性。
- 3) 处于网络运营商范围之外的支撑设备的性能（例如用户设备、主机）。
- 4) 用户任务，以及用户与该应用互动的程度。

为确保用户满意，网络设计者要考虑这些因素。数据应用性能的模型应该尽可能多地包括这些因素。一旦已经评估了应用性能，则能够应用感知模型来解释得到的端对端性能水平。

本建议书假设读者将至少能够对以上每个关键因素提供一定程度的详细情况，并因而能采用工具（其中很多在此确定）框架来评估端对端性能。

为数据应用评估IP网络端对端性能

1 范围与应用

本建议书涉及对在IP网络上运行的应用的端对端性能评估程序，采用：

- 所关心的IP网络的性能，基于相关测量或网络模拟的结果。
- 所关心的应用的规范，根据其带有特定选项的控制协议，或采用网络性能和客户设施性能作为输入，并生成应用性能关键尺度（例如文件下载时间）作为结果的应用模型。
- 一个专用于所关心的应用的感知模型，将应用性能解释为对由典型用户大众所体验到的质量评估。

图1显示了进行一个端对端性能评估的一般程序。

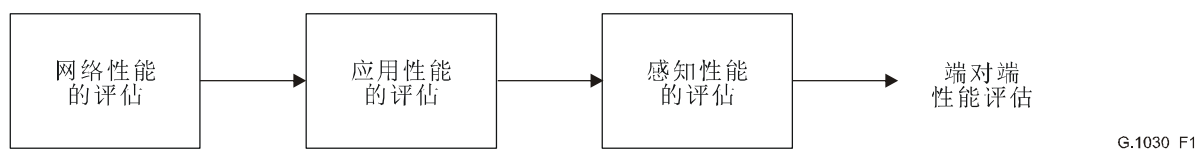


图1/G.1030 – 获得端对端性能评估的程序

请注意，获得网络和应用性能的步骤在一些情况下可以合并，例如一个仿真过程提供一个特殊进程或一组进程的性能测量方法的情况。

1.1 网络性能的评价

可以根据ITU-T Y. 1540建议书和其他相关标准（例如由IETF IP性能尺度工作组制定的RFC）中定义的包传送性能参数来评价网络性能。除了通常的单向传送性能尺度外，对应用性能最直截了当的评价有时可能还是往返尺度。

有两种主要的网络性能信息来源：测量与模拟。

网络测量允许评价者将网络作为一个黑盒子来对待，并且生成可以对模拟环节的其余步骤有用的信息。但是，对测量的设计有一些重要的考虑，包括：

- 1) 在一些情况下要求发送约束与所关心的应用相匹配。例如，TCP的流控制要适应网络条件并有可能要在瓶颈处排队，在其他情况下测得的时延之外增加了时延。
- 2) 非测量负荷必须在大小和特性上与性能评估可以适用的条件相似。在未加负载的网络上进行的测量不是特别有用。

网络模拟能够在网络还没有完全建成时（例如，节点和链路已经到位，但没有运行关键协议），或者测量的关键考虑无法实现时提供所需要的性能特性。有许多模拟工具可以选择，包括商业产品和公共领域研究工具。模拟工具要求对所关心网络的专门技术和信息要达到充实的程度，以便有效地运用它们。作为对此相当精确网络模拟的一个替代，附录一提供了可以用于评估IP网络中端对端性能的简化方法。但是，此方法的准确性将很大程度上取决于所提供信息的准确程度。

1.2 应用性能的评价

应用模型将网络性能评估和描述应用设备性能的信息作为输入，并产生一个或多个应用性能关键尺度作为输出。

在ITU-T G.1040建议书中已经确定了一个这样的应用性能模型，用于典型的信用卡和其他销售终端（POS）事物的对话型包交换。

对采用TCP可靠字节传送业务的长持续时间文件传送，只要输入正确，附录一中所描述的模型可提供准确的结果。

1.3 感知模型

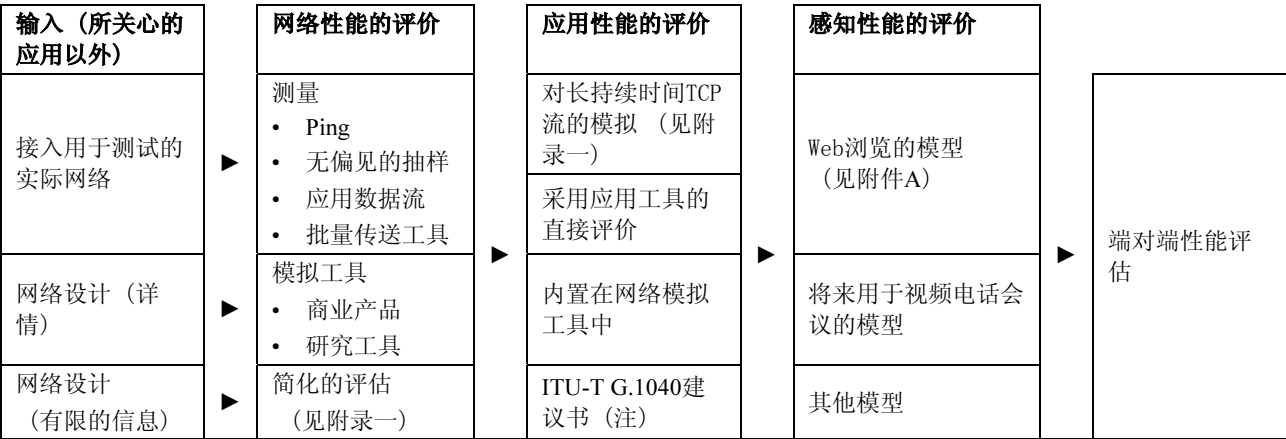
本建议书给出了用于获得期望的端对端性能评价的感知模型。

附件A为Web浏览确定了一个模型。

当有另外模型可用时，它们将作为附件纳入本建议书。

1.4 端对端性能评价程序中模型的框架

图2显示了在评估IP网络中应用的端对端性能过程中的各种可选择方式。该图表明，有许多选择可用于完成这个过程，尽管在实际中评价者必须将与一个端对端评估的目标一致的选项组合起来（且各选项间相互一致）。



注 – ITU-T G.1040建议书给出了事务应用性能中网络所占份额，未提供完全的应用性能评估，但它仍给出了网络性能的面向应用的视图。

图2/G.1030 – 用于开展端对端性能评估的框架

2 参考文献

下列ITU-T建议书和其他参考文献的条款，在本建议书中的引用而构成本建议书的条款。在出版时，所指出的版本是有效的。所有的建议书和其它参考文献均会得到修订，本建议书的使用者应查证是否有可能使用下列建议书或其它参考文献的最新版本。当前有效的ITU-T建议书清单定期出版。本建议书引用的文件自成一体时不具备建议书的地位。

- ITU-T Recommendation G.1010 (2001), *End-user multimedia QoS categories*.
- ITU-T Recommendation G.1040 (2006), *Network contribution to transaction time*.
- ITU-T Recommendation Y.1540 (2002), *Internet protocol data communication service – IP packet transfer and availability performance parameters*.

3 缩写词

IP	网际协议
QoS	服务质量
TCP	传输控制协议

4 最终用户QoS的评价

在对端对端物理性能进行评价之后，网络规划者绝对有必要在考虑感知方面问题的情况下评价最终用户的QoS。

ITU-T G.1010建议书规定了数据应用以及语音和视频应用的QoS类别。此外，附件A为评估用户对Web浏览应用的感知质量规定了一个评价模型。这使得有可能以测得的/计划中的端对端时延数值为基础按照客户的意见进行诊断评价。

附件 A

Web浏览应用的评价模型

A.1 范围

本附件规定了一个模型，用于将从网络中测得的或从HTTP事务处理时间算得的响应和下载时间映射为Web浏览进程的感知质量。该模型是基于处理Web浏览进程中响应和下载时间的实验[1], [2]。该模型的范围目前还仅限于包含两个步骤的Web浏览进程，第一步发出搜索请求，第二步显示页面结果。还为单个定时事件提供了简单的扩展，它模拟了等待单个页面的影响。

A.2 引言

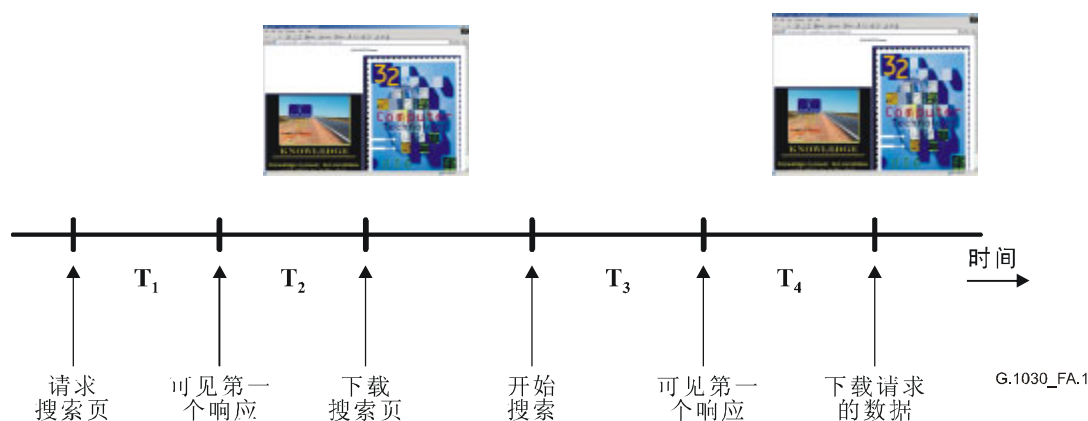
在Web浏览感知质量模拟中的一个重要观察结果是预期的最大进程时间在感知质量中占据主导地位这样一个事实。如果一个人希望进程时间为100秒钟，则一个10秒钟进程的感知质量将远高于一个人希望进程时间为1秒钟的情况。因此，该模型采用了随内容而定的方式，使用了三个时间尺度，6、15和60秒钟，分别对应于快速、中速和慢速网络内容。

一般而言，与响应时间相关的质量感知能够根据以下三个感知范围来分类 [3]:

- 1) **瞬间体验:** 0.1秒钟大约是感觉系统瞬间进行响应的极限，它是会话业务的一个重要极限（例如聊天）。
- 2) **无中断体验:** 1.0秒钟大约是保持用户思维不被中断的极限，即使该用户确实失去服务是一直运行的感觉，它是互动服务的一个重要极限（例如游戏）。
- 3) **失去注意力:** 10秒钟是将用户注意力保持集中在一个对话上的大致极限。对更长的延时，用户会希望在等待计算机完成操作时执行其他任务，所以应该给用户一个反馈信息，通知他们预计什么时候计算机可以完成操作。如果响应时间有可能在大范围内变化时，延时期间的反馈就特别重要，因为用户不知道将会发生什么。

对于下载时间，受试者对质量的判定倾向于随预期的下载时间而变[4]。当受试者得知预期的下载时间时，他们会愿意接受较长的下载时间。

本建议书中的模型描述对于一定的网络和系统配置，web浏览进程中不同的响应和下载时间与相应的给定最大进程时间的Web浏览进程的感知Web浏览质量之间的关系。该模型可适用于范围广泛的网络和系统配置，以及各种类型用户的Web浏览服务。作为该模型基础的主观实验尽可能逼真地模仿了真实生活中的Web浏览体验。在模型的开发中，采用了三个主观Web浏览实验，时间尺度大约为6、15和60秒钟，分别表示快速、中速和慢速网络内容。在每个进程中，一个受试者首先更新一个搜索页面，然后得到一个显示搜索结果的页面。图A.1中显示了这样一个进程的时间过程。前面的两个时间间隔 T_1 和 T_2 给出了搜索页面的非互动响应及下载时间。后面两个时间间隔 T_3 和 T_4 给出了结果页面的互动响应和下载时间。



T_1 为非互动时间，并且采用Java脚本处理。 T_2 为非互动下载时间，并且采用网络操作器来处理。 T_3 和 T_4 为互动部分的对应时间。 $T_1+T_2+T_3+T_4$ 之和代表该进程时间。

图A.1/G.1030 – 实验计时器

由于已知训练有素的专家与所谓缺少经验的用户的行为之间存在差别，在开发模型时对这两组人做了区分。将这两个组区分开来使得我们能够开发一个对广大用户人群预测Web浏览质量的模型。

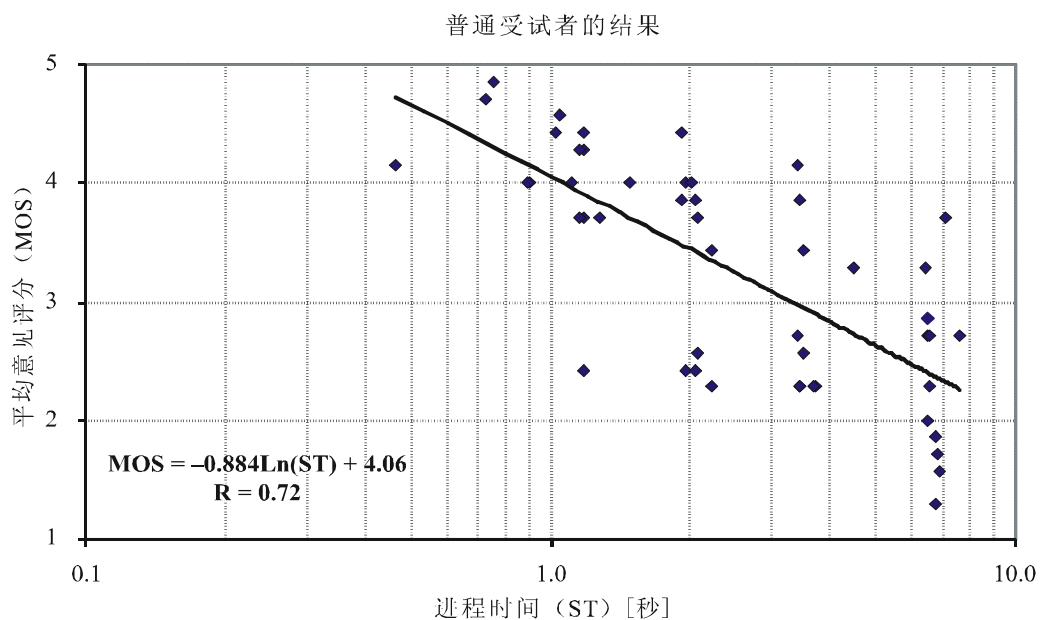
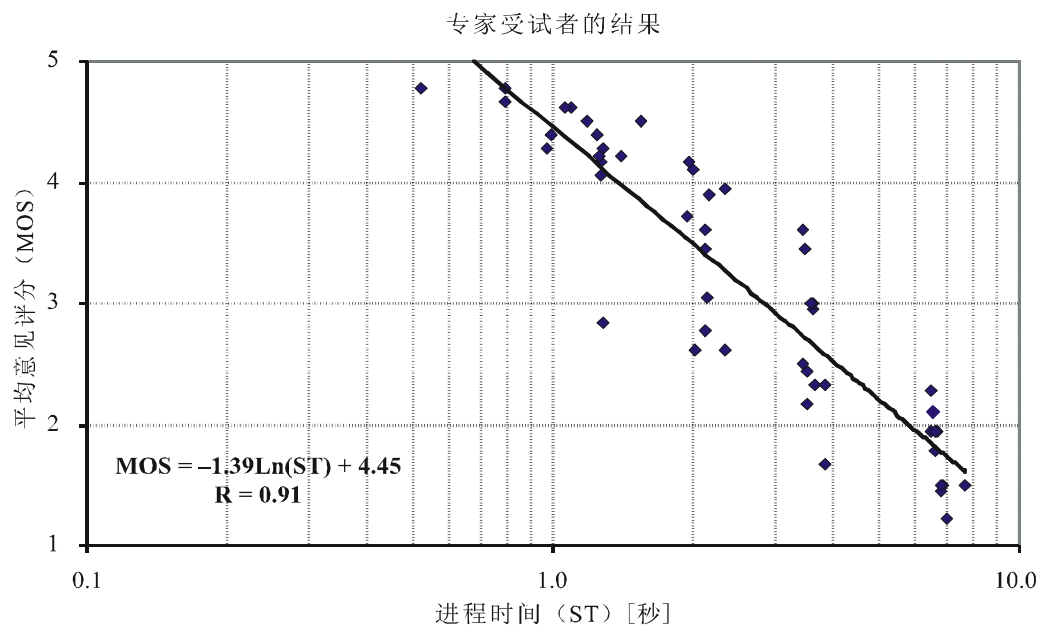
A.3 主观Web浏览质量实验与结果

在这些实验中，进而在该模型中，采用了ITU-T的绝对类别评分尺度[5]（五分制，此处5分：优秀，4分：良好，3分：尚可，2分：低劣，1分：差）。在每个实验中，进行了49个进程，每个进程包括：

- 请求、更新和显示搜索页。
- 在此页上键入并提交一个搜索项。
- 更新并显示结果页。

为了得到一致的数据，每个进程采用了完全相同的结果页，并且要求受试者在每个进程中键入相同的搜索查询。从受试者的观点看，搜索引擎首先得要找到结果页，然后得要将其下载。对49个进程中的每一个，配置了不同的 T_1 到 T_4 组合，改变 $T_1+T_2+T_3+T_4$ 总和，即进程时间，从0到对这个实验组的时间尺度。

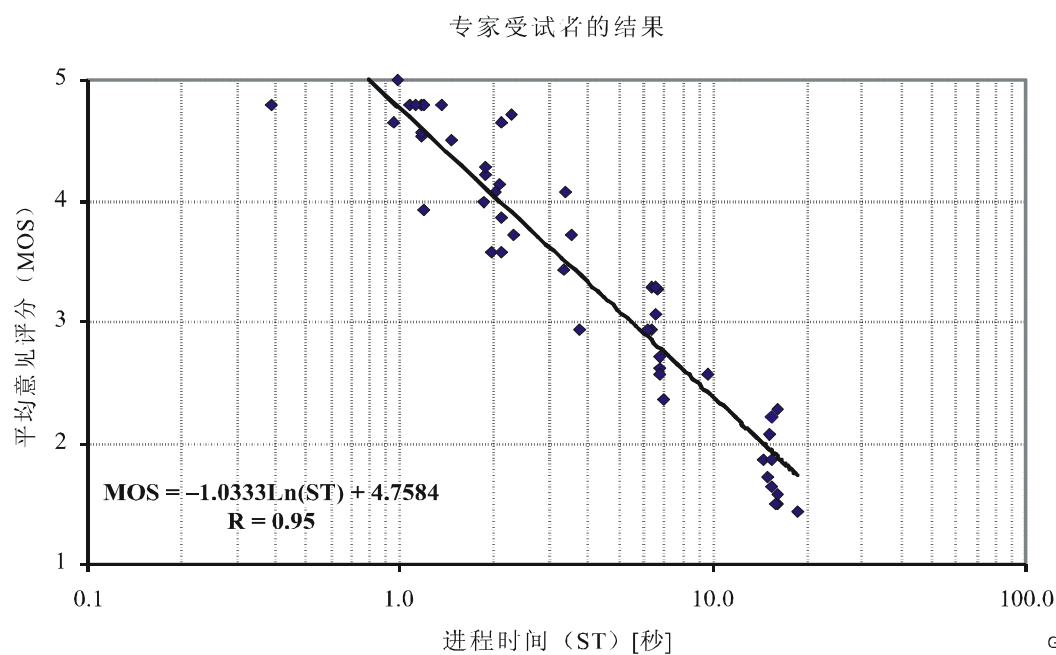
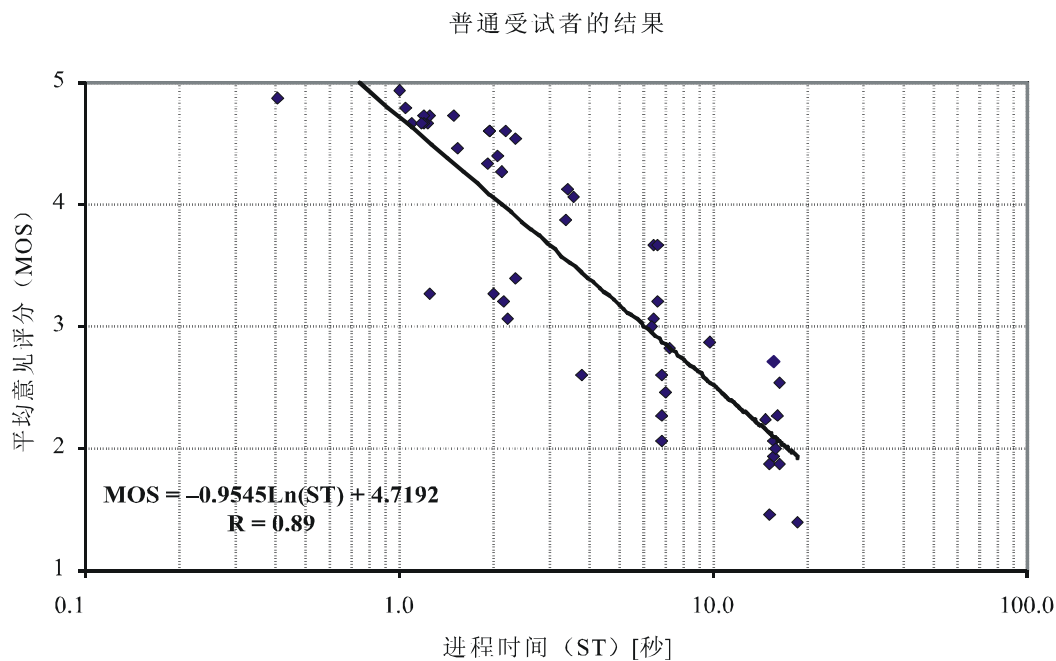
总的结果在图A.2-A.4中给出，图中对所有三个实验采用进程时间（即 $T_1+T_2+T_3+T_4$ ）对应于MOS值画图。对长持续时间的内容，请见图A.4，结果对普通受试者和专家受试者基本是相同的，而且MOS是对所有人群计算的。对两个较短持续时间的内容，见图A.2和A.3，普通受试者和专家受试者表现不同，而且进程时间和感知质量之间的关联明显不同，并因此而分别给出。所有结果显示了相同的行为，感知质量按进程时间的对数线性下降。在长持续时间实验内容中的关联性（ >0.9 ）高得足以对普通受试者和专家受试者进行可靠的质量预测。通常，用于质量感知的心理物理建模的相关性需要达到0.9以上[6]。结果还显示，在普通受试者的6秒钟实验中，进程时间与感知质量之间的相关性低得太多（0.72），不足以使用仅仅基于进程时间的简单模型。



G.1030_FA.2

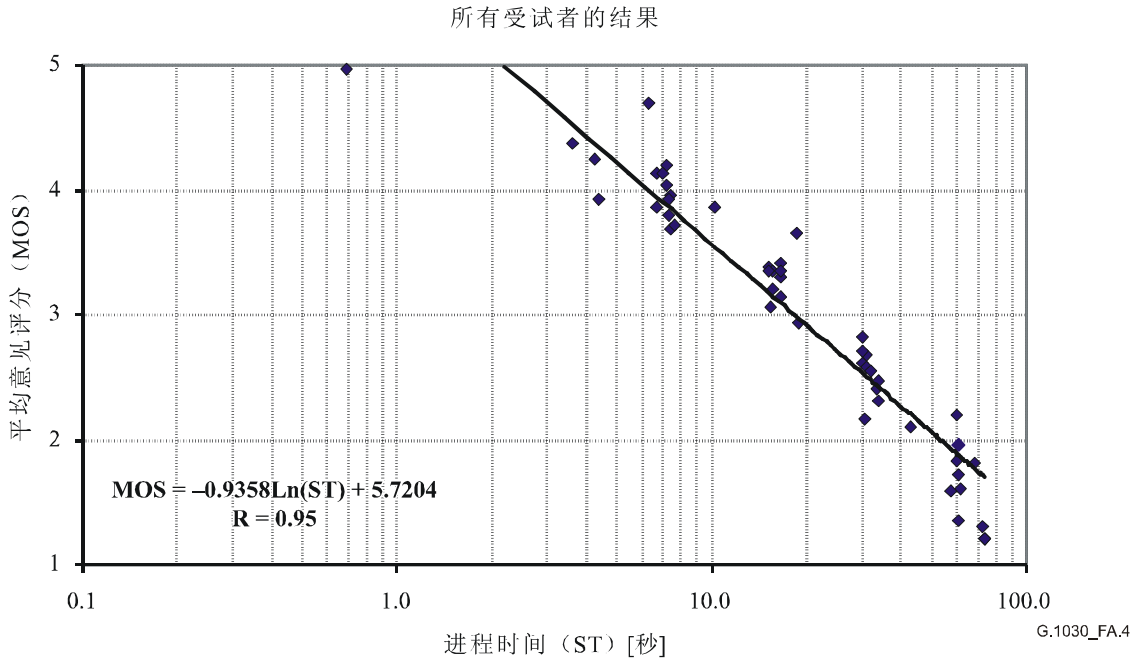
时间尺度为6秒钟时对（7）个普通受试者和（18）个专家受试者的结果。对普通受试者，相关性低得不足以实现准确的MOS预测。对专家受试者，可以从0.67和12秒钟之间进程时间的对数插值预测MOS。

图A.2/G.1030 – 受试者在时间尺度为6秒钟的结果



时间尺度为15秒钟时对（15）个普通受试者和（14）个专家受试者的结果。二者显示了相同的表现。对普通受试者，相关性稍微低于可以进行准确的MOS预测。对专家受试者，可以从0.79和38秒钟之间进程时间的对数插值预测MOS。

图A.3/G.1030 – 受试者在时间尺度为15秒钟的结果



时间尺度为60秒钟时对所有（12+17=29）个受试者的结果。可以从2.16和155秒钟之间进程时间的对数插值预测MOS。

图A.4/G.1030 –受试者在时间尺度为60秒钟的结果

A.4 模拟结果

A.4.1 60秒钟内容

对60秒钟内容，进程时间与主观质量之间的关联性对普通受试者和专家受试者都非常好（0.95），并且从进程时间到主观质量的映射能够直接用于客观模型中。回归显示，大约2秒钟以下的进程时间引起最大MOS值5（见图A.4），而最小MOS值1是在进程时间大于实际实验中最长进程时间时达到。对长持续时间内容的从进程时间到Web浏览质量的一般映射是通过定义一个最小（*Min*）和一个最大（*Max*）进程时间，然后在这些进程时间极值之间采用对数插值来构成的。如果写出 $MOS = a - b \cdot \ln(\text{SessionTime})$ ，对 $\text{SessionTime} = \text{Min}$ 填入 $MOS = 5$ ，对 $\text{SessionTime} = \text{Max}$ 填入 $MOS = 1$ ，我们就得到 *Min* 和 *Max* 之间的进程时间：

$$MOS = \frac{4}{\ln(\text{Min} / \text{Max})} \cdot \left(\ln(\text{SessionTime}) - \ln(\text{Min}) \right) + 5 \quad (1)$$

对长持续时间实验（见图A.4），回归为：

$$MOS = 5.72 - 0.936 \cdot \ln(\text{SessionTime}), \quad (2)$$

介于MOS 1.0和5.0之间。

A.4.2 6秒钟和15秒钟内容

对6和15秒钟实验内容，进程时间与主观质量之间的关联性远低于60秒钟内容情况，一个用于对普通受试者和专家用户主观质量预测的更先进的模型是通过采用这样一个观念来构成的，即对较短的进程时间，最后的下载时间（在我们实验中的 T_4 ）对最终感知Web浏览质量比其他响应和下载时间（在我们实验中的 T_1 、 T_2 、 T_3 ）具有更大的影响。表A.1给出了加权因子，为了得到具有与主观确定MOS值的最高关联性的量值，必须采用这些加权因子来对 T_1 到 T_4 进行加权。这个量值，即加权的进程时间，

$$WeightedST = WT1 \cdot T1 + WT2 \cdot T2 + WT3 \cdot T3 + WT4 \cdot T4$$

能够通过采用与公式（1）中所采用的相同的最小和最大进程时间之间的对数插值来映射得到MOS值：

$$MOS = \frac{4}{\ln(Min / Max)} \cdot (\ln(WeightedST) - \ln(Min)) + 5 \quad (3)$$

表A.1显示，对较短持续时间内容，最后下载时间的影响比其他下载和响应时间影响的二倍还大。它还显示出普通受试者和专家受试者极为不同的表现，对普通受试者的最佳权重对最后下载时间的影响比对专家受试者最佳权重的更大。对普通受试者，长下载时间的影响比其他下载和响应时间的影响的4倍还大。对专家受试者，它大约是2倍，而总的最佳权重显示出的影响大约是3倍。

**表A.1/G.1030 –客观定时和主观OMS结果之间具有相关模型关联性时的
T₁、T₂、T₃和T₄的最佳模型加权**

	WT1	WT2	WT3	WT4	最小	最大	相关性
6秒 专家	0.56	0.84	0.80	1.80			0.97
6秒 普通	0.37	0.40	0.60	2.63			0.93
6秒 总体	0.47	0.60	0.71	2.22	0.62	13.5	0.95
15秒 专家	0.63	0.77	1.11	1.49			0.98
15秒 普通	0.48	0.70	0.88	1.95			0.96
15秒 总体	0.54	0.72	0.98	1.76	0.81	39	0.97
60秒 专家	0.84	0.77	1.22	1.18			0.99
60秒 普通	0.64	1.01	1.12	1.24			0.98
60秒 总体	0.73	0.90	1.16	1.22	2.22	151	0.98
60秒 总体，未加权，见图A.4	1.00	1.00	1.00	1.00	2.16	155	0.95

注 – 为了能够将一般进程时间（T₁+T₂+T₃+T₄）与加权进程时间进行比较，将加权系数的总合归一化为4.0。在标准模化模型中使用的加权以及公式（3）中所用的最小和最大时间以加粗字体显示。对>60秒钟的长进程时间内容可以采用不加权的简单模型（加粗斜体字）。

表A.1还显示了对中等持续时间内容的结果，对普通受试者和专家受试者的加权因子以及总的加权因子处于对短的和长的持续时间内容实验的加权因子之间。这显示了加权方式的有效性，因此使得有可能在不同实验进程内容时间（即，时间尺度6、15和60秒钟）之间插值来得到对其他内容时间的加权。

A.4.3 总结

本建议书为60秒钟持续时间内容标准化了一个简单模型，它使得可以采用公式（2）评价最大进程时间大约为155秒钟的Web浏览进程。本建议书进一步标准化了三个先进的模型，它们采用来自表A.1中的最佳总权重，结合以下从加权进程时间到以平均主观评分表示的感知浏览质量的映射：

$$MOS = 4.38 - 1.30 \cdot \ln(WeightedSessionTime) \quad (4)$$

对短持续时间进程，处于MOS 1.0和5.0之间。

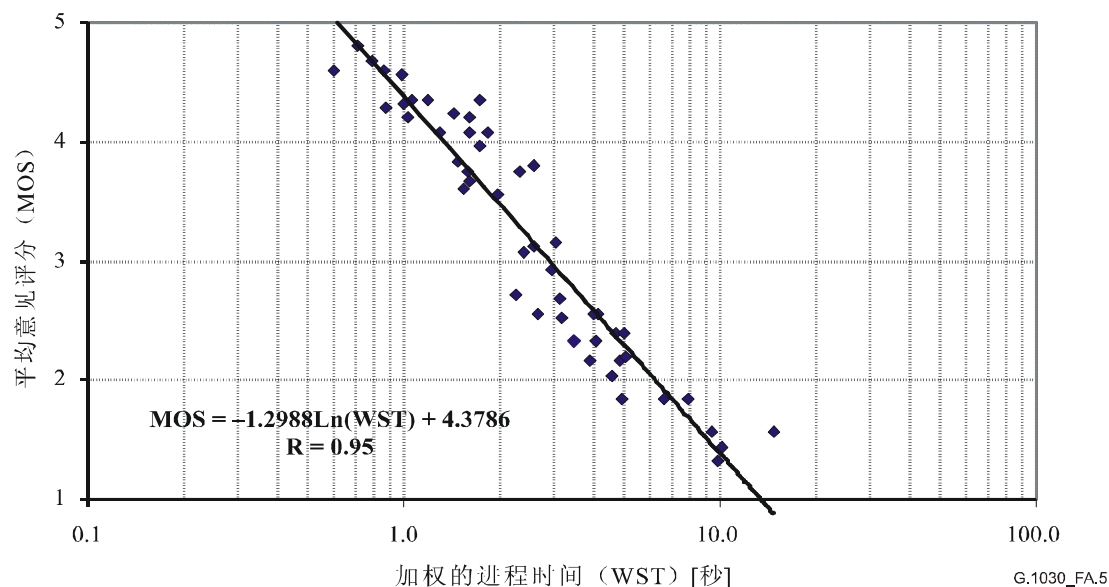
$$MOS = 4.79 - 1.03 \cdot \ln(WeightedSessionTime) \quad (5)$$

对中等持续时间进程，处于MOS 1.0和5.0之间。

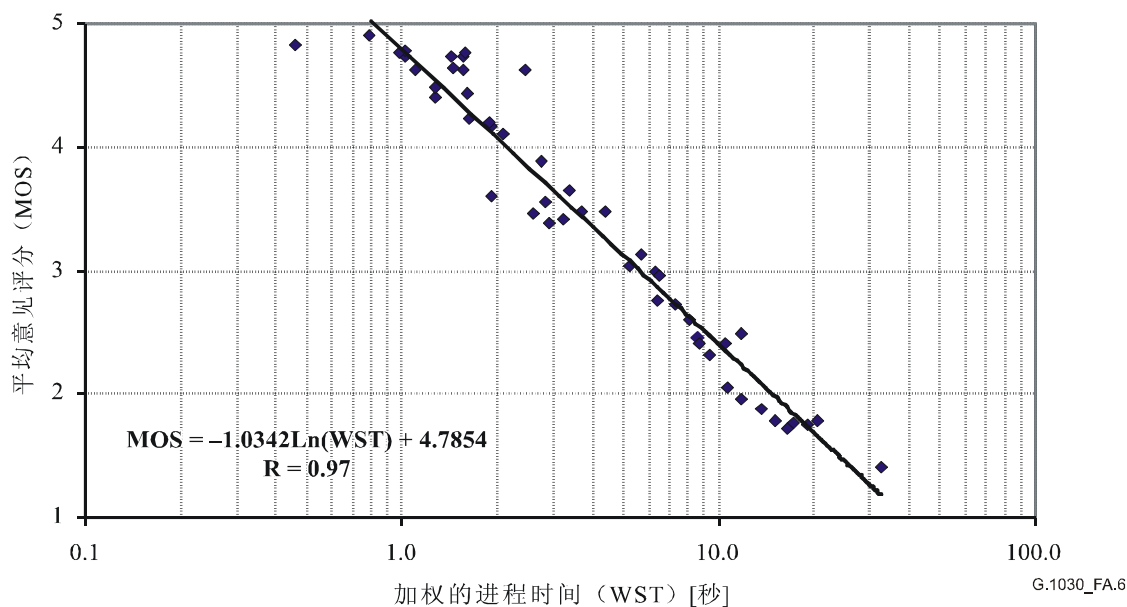
$$MOS = 5.76 - 0.948 \cdot \ln(WeightedSessionTime) \quad (6)$$

对长持续时间进程，处于MOS 1.0和5.0之间。

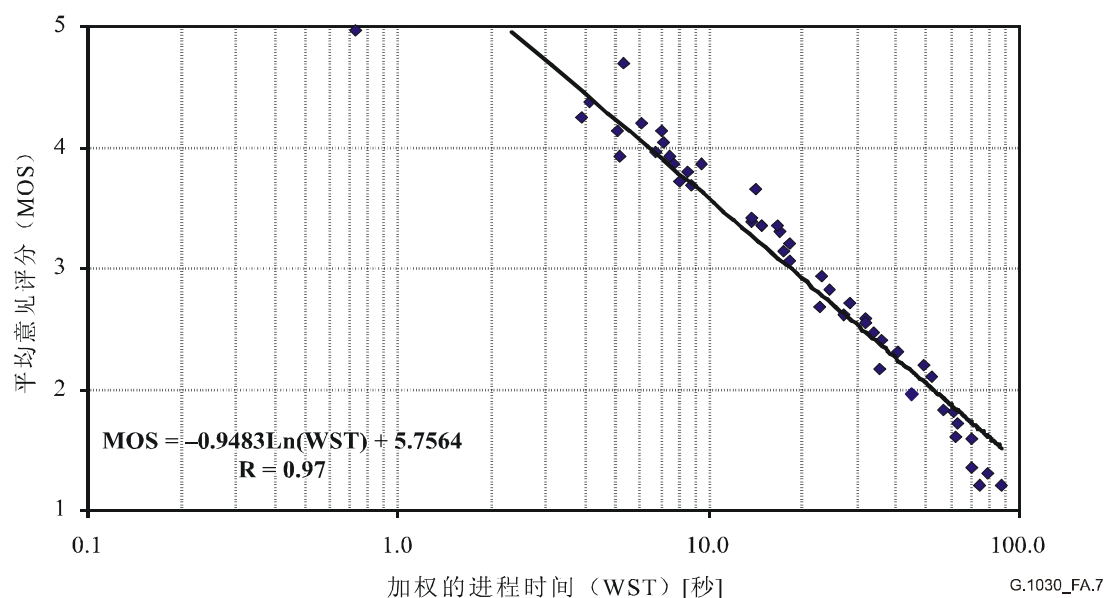
在图A.5-A.7种给出了来自公式（4）、（5）和（6）回归适配的结果。



图A.5/G.1030 – 所有时间尺度为6秒钟的受试者的模型与数据的对应



图A.6/G.1030 – 所有时间尺度为15秒钟的受试者的模型与数据的对应



图A.7/G.1030 –所有时间尺度为60秒钟的受试者的模型与数据的对应

A.5 一页Web浏览进程和单个定时事件的感知质量（非标称化的）

基于作为推导公式（1）和（3）出发点的实验数据，我们还推导出包含单个Web网页的Web浏览进程的进程时间与感知质量之间的关系。此外，我们还对下载时间总是等于0的单个页面的进程感兴趣。此情形发生于更新的信息在一定等待时间后即时呈现给用户时。这个最后的关系将单个定时事件以MOS评分映射到主观感知质量。

一页映射推导的出发点是观察到质量随着最小和最大进程时间之间进程时间的对数线性下降，最小和最大进程时间分别对应着MOS评分5.0和1.0（见图A.2-A.7）。从已经在图A.2-A.4中给出的二页浏览数据我们看到，在三个所采用的主观实验内容中的最小进程时间在大约0.7与2.2之间变化，并且随着实验中的最大持续时间增加时而增加（见表A.2中的结果总结）。

表A.2/G.1030 – 在二页浏览实验中未加权的最小和最大时间

	最小（秒）	最大（秒）
6秒专家	0.67	12
15秒专家	0.79	38
60秒总体	2.16	155

我们能够根据这些结果定义一个最小进程时间（*Min*），它得到一个5.0的MOS评分， $Min = 0.011Max + 0.47$ ，*Max*表示期望发生的最大进程时间。这使得可以实现对二页实验在任何期望二页Web浏览进程最大持续时间（*Max*）情况下从进程时间到MOS评分的一般映射：

$$MOS_{2-page} = \frac{4}{\ln((0.011Max + 0.47) / Max)} \cdot (\ln(SessionTime) - \ln(0.011Max + 0.47)) + 5 \quad (7)$$

从已经给出的实验数据，我们可以估计这个关系很可能适用于大约在10到200秒之间的所有二页Web浏览进程。

对只有一个响应时间 T_1 以及一个下载时间 T_2 （见图A.1）的任意一页进程，最高质量是在大约为二页进程中所得数值一半的最小数值时感知到。这导致了最小进程时间的定义为 $Min = 0.005Max + 0.24$ ，同时对MOS值的映射由公式（8）给出：

$$MOS_{1-page} = \frac{4}{\ln((0.005Max + 0.24) / Max)} \cdot (\ln(SessionTime) - \ln(0.005Max + 0.24)) + 5 \quad (8)$$

从已经给出的实验数据，我们能够估计此关系很可能适用于大约5到100秒之间的所有一页Web浏览进程。

类似地，对下载时间总是等于0的任意一页进程，或对于任何单个定时事件，我们得到：

$$MOS_{single\ timing\ event} = \frac{4}{\ln((0.003Max + 0.12) / Max)} \cdot (\ln(SessionTime) - \ln(0.003Max + 0.12)) + 5 \quad (9)$$

从已经给出的实验数据，我们能够估计这个关系很可能适用于大约3到50秒之间的所有单个时间事件。请注意，对这样一个单个互动，最小时间等于0.12秒，对应于瞬间感知门限[3]。

应该注意，以上对一页Web浏览进程和单个定时事件的感知质量应该通过实际实验来验证。

A.6 附件A的参考文献

- [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*.

附录一

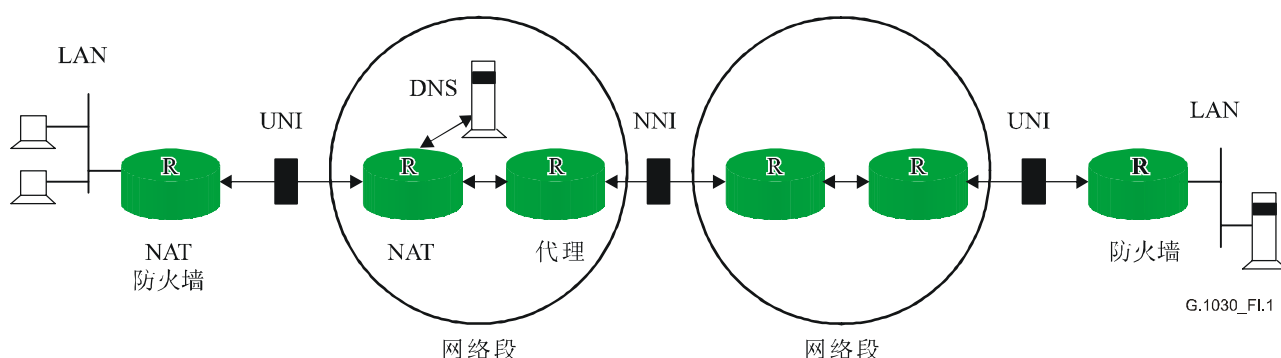
采用有限信息的网络性能评估

I.1 引言

本附录提供当有限信息可用时可以用于评估在一个IP网络中端对端性能的简化方法的信息。采用这些方法推导出的评估的准确性高度取决于输入信息的质量。在进一步研究与评价之后，本附录中的材料可能会改变。

I.2 参考连接

有必要解决参考连接中的防火墙、代理服务器（为了Web和/或性能的增强）、网络地址翻译器（NAT）以及负载均衡开关，因为这些设备今天存在于很多家庭、企业和受管理IP网络之中。这些“中间盒子”提供很多的功能，但是消除了端对端透明性这个所希望的IP网络结构特点。图I.1 显示了适当的参考连接。



此扩展允许信令协议性能的评价，例如采用与其他IP应用相同总体框架的SIP。

I.3 包传送性能数值的级联

引入中间盒子意味着在端对端路径中将有更多分立的段落，导致需要准确的累积公式。

I.3.1 时延

对分立段落的平均时延是累加的。但是，可能要注意，平均时延表示在包传输中经常会观察到的时延偏差分布的重心。当通过一个网络段落的包传输时延（或在一个主机中的处理时间）由该平均值来表示时，来自时延分布的单个抽样被其期望值代替。时延的时间平均值在此处是恰当的，因为每个处理将对整个处理间隔内很多不同瞬间处的潜在时延分布进行抽样。

I.3.2 丢包

端对端丢包率（IPLR）性能本质上是UNI-UNI性能。可以假设终端设备和主机的丢包概率可忽略不计。

所建议的IPLR级联方法是求通过 n 个网络段落成功包传送概率的反转概率，如下所示：

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

此公式取决于条件概率理论并且假设在每个网络段落中的丢包概率相互独立。对一条具有两个丢包概率分别为 p_A 和 p_B 的网络A和B的UNI-UNI路径：

$$\begin{aligned} \text{概率}\{\text{在两个段落都成功}\} &= \text{概率}\{\text{在B上成功}|\text{在A上成功}\} \times \text{概率}\{\text{在A上成功}\} \\ &= (1 - p_B) \times (1 - p_A) \end{aligned}$$

对各个网络（ p_A 和 p_B ）根据经验推导出来的丢包概率等于对任何在前面的网络上成功的条件概率 – 它们需要一个包进入事件来开始测量操作。

I.3.2.1 突发丢包的叠加

留待将来研究。

I.3.3 时延偏差

传统IP和许多数据应用对时延偏差的敏感性要低于要求连续输出率的应用（同步或实时应用，例如VoIP）。出于这个原因，传统IP应用有时因为它们“弹性”分组流而被划分为一类，它们可以在传送期间被压缩或拉长而影响很小。当VoIP包间隔在传送中发生改变时，应用会要求附加缓存器来恢复连续输出能力，并且缓存器本身对这些“非弹性”分组流就是一个时延的根源。

很难以一个准确的方法将各种网络段落和处理器的时延偏差叠加起来，因为很少知道完整的时延分布。但是，在8/Y.1541节中给出了一个合理的近似方法。

I.4 瓶颈带宽

在计算端对端数据传送时间时采用接入带宽作为限制因素能够导致对吞吐量过度乐观的看法。

I.4.1 关键考虑

通过将以下考虑包括在模型当中，能够得到一个对固定长度文件或信息的实际数据传送时间更准确的近似：

- 1) 必须给出传送方向。对很多新技术，接入速率是非对称的，例如DSL和电缆调制解调器，而且差别可以达到10倍或者更大。
- 2) 采用真实瓶颈带宽。接入速率并非必定是瓶颈。主机之一可以扼制传输速率（不能将足够的数据送到TCP处理，或者带有窗口限制），主机TCP参数能够限制吞吐量，而诸如提高性能代理服务器这样的中间盒子可能会改变TCP参数来改善吞吐量，等等。
- 3) 事务可能会遇到对共享资源的竞争。采用由设计目标隐含的有效瓶颈带宽来在一个链路或其他共享资源上支持若干数量的同时用户。
- 4) 瓶颈带宽可以是随时间变化的。带宽限制将产生一个数据传送时间范围。

- 5) 必须在瓶颈点计算开销。在IP以下的每一层都增加一些信头开销。假设一个有效载荷大小并且计算开销百分比，这将进一步增加数据传送时间。
- 6) TCP达不到等于瓶颈带宽的传送能力。Mathis的TCP公式对稳态传送能力给出了很好的近似，然而以上因素可以具有压倒性的重要性。

这些项中的大部分需要对模型的用户或网络规划者部分有更多的准备。在I.4.2给出了TCP能力近似的细节。

I.4.2 TCP能力

传统IP应用中的绝大多数采用TCP（传输控制协议 - Transmission Control Protocol）的可靠字节流传送服务。一个对TCP流控行为的模型被描述为[TCPCOn]：

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

其中：

BW 是每单位时间（循环时间）传送的数据

MSS 是TCP最大分段大小

RTT 是平均往返时间

C 是一个考虑随机/周期丢包和ACK策略效应的常数

p 是丢包概率

即使对最近的TCP增强，例如SACK和TCPReno，此公式应该被视为是对信息流的一个上限。假设，TCP连接持续时间足够长使在拥塞避免状态中达到平衡。当在数据传送中总的字节很小时，TCP流控可能达不到平衡，并且慢启动特性主导了计算。在[Padhye] 可以找到一个对TCPReno能力更为灵活的关系，包括对最大窗口大小的限制效应。

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

其中：

$B(p)$ 是TCP吞吐量的近似模型[包/秒]

$W_{\max}$ 是接收机最大窗口缓存器大小[包]

RTT 是往返时间[秒]

b 是由一个接收的ACK确认的包数量

p 是一个包丢失的概率

T_0 是重新传送一个未确认（丢失）包的超时时间[秒]

TCP的流控企图增加其发送速率（窗口大小）直到它遇到拥塞（或一个瓶颈链路），并且它从包丢失来对此做出推断。因此，一些包丢失对探索瓶颈是固有的。当丢包发生时将窗口减半，然后每次将窗口增加一个包直到发生另外一个丢包，这样的避免拥塞流控有效地将吞吐量限制在窗口峰值大小或发送速率的75%。我们能够重新安排此公式来解决仅仅由于TCP瓶颈探测造成的丢包。

$$\sqrt{p_{\text{Probing}}} \approx \frac{MSS}{RTT} \frac{C}{0.75 \times BW(\text{bottleneck_link})}$$

表I.1列出了与详细瓶颈链路速度相关的固有丢包。

表I.1/G.1030 – 由于瓶颈探索造成的TCP丢包

C	MSS	RTT	BW (链路)	75% BW	Delay*BW	窗口	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

在表I.1中可以注意到几点：

- 当瓶颈链路带宽为10 Mbit/s（以太网或更高）时，固有丢包非常低。如果网络丢包为 10^{-4} 或更高，将往往是按照[TCPCon]减少吞吐量。
- 对给定的参数（80 ms RTT，12 000比特（1500字节）MSS，等），当单个TCP流探测瓶颈时，在一个 T_1 链路上大约100个包中有一个包丢失。 10^{-3} 或更低的网络丢包对产生的吞吐量几乎是不合理的。
- 显示了路径的Delay*BW的积和最佳窗口大小（product/MSS），并注意，在窗口少于一且计算的丢包率无意义（>1）处，有两个链路带宽（128 kbit/s和64 kbit/s）。在这些情况下TCP调谐是有保证的，所以我们将MSS减少到2048 bits并且产生一个更合理的结果。
- 请注意，在这里未曾涉及到由于低层开销造成的带宽减少。

评估的网络丢包率将和探测丢包率结合如下：

$$p_{\text{Total}} = 1 - \{(1 - p_{\text{Network}}) \times (1 - p_{\text{Probing}})\}$$

当网络和探测丢包率幅度是相同数量级时，则将它们合并起来计算带宽有点悲观，因为一些探测丢包并不会发生。

最后，可能要注意，除了提到在某些情况下参数能够限制吞吐量，这里并没有详细涉及TCP参数调谐。如果最大窗口大小不足以满足往返路径，则吞吐量将被限制为窗口 $\times$ MSS/RTT [Padhye]。很多TCP连接在其寿命中从来没有经历过丢包，这归咎于小窗口和TCP填充位于瓶颈之前排队的倾向（并且对一些包增加RTT – 这就是为什么采用平均RTT）。

I.5 握手时间

包交换可以分为两个阶段：

- 1) 握手；及

- 2) 数据传送（一些应用不具有这一阶段，例如信用卡的销售终端（POS）和VoIP信令。作为一个简化，我们将在此阶段包括最后的包交换包来完成连接。传统的 3×3 矩阵将断开连接当作分开的阶段对待）。

如果一个包或其响应在初始握手期间丢失，发送者通常在重新发送该信息之前等待一定的时间。此等待时间，或重新传送超时，能够是一到三秒数量级，当发生丢包时将明显延长初始握手时间。

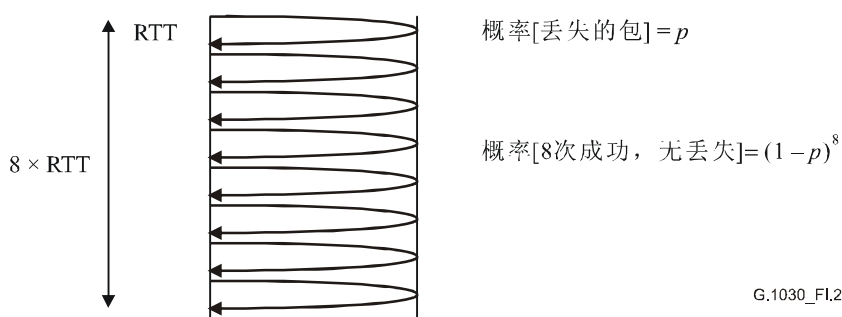
如果将所有与网络相关的时间（包括那些由丢包引起的）与主机处理时间分离开，存在着二种与可分离管理责任（网络和CPE/主机）相结合的尺度。

这个分解对初始握手阶段是相当直截了当的。包传送时间和由于丢包的超时归咎于网络。

对数据传送时间，给出时间并显示限制因素是否是网络丢包、网络时延或主机处理/设置可以是足够的。

I.5.1 握手期间丢包的影响

IP网络对握手时间的贡献能如以下显示的来决定。首先，选取有8个信息和响应必须要交换，并且所有包传送都是成功的情况。这种情况的概率显示如下：



图I.2/G.1030 – 带有8个RT返回及没有丢包的握手

这是一个简单案例，但它代表了IP网络对往返丢包概率 p 足够低时几乎所有用户体验的贡献。我们注意到，往返丢包概率为：

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

完成所有8个交换的时间是8倍的平均往返时间（RTT），加上远程主机处理时间（HPT），这个总值就是握手时间。这些成分可以分开计算，因为：

$$Handshake_Time = NCTT + Total_HPT$$

此处，NCTT是网络对事务时间的贡献，如在ITU-T G.1040建议书中所定义。

请注意，POP3握手的例子具有8个请求/响应交换。如果我们假设 $p = 10^{-3}$ ， $RTT + HPT = 0.080$ 秒，重新发送超时（RTO）= 1秒，并且有35万次事务尝试，我们就得到如表I.2中所示的对每个丢包情况的概率。

表I.2/G.1030 – 对带有0到3个丢包情况的握手时间

丢包	握手时间 [秒]	发生概率	事务
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

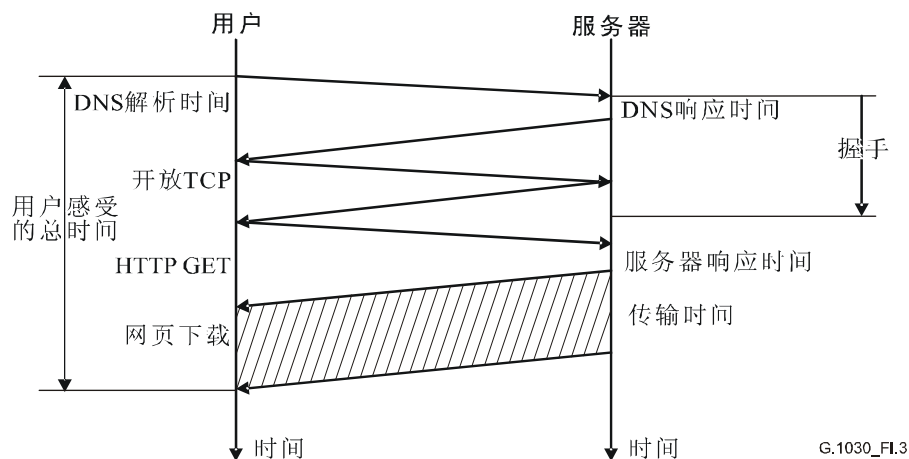
这个例子显示了IP网络对一个带有8个RT交换的POP3事务的平均、第95百分点和第99百分点握手时间的贡献少于1秒。

I.5.2 数据传送期间丢包的影响

丢包需要重新传送，而且有了TCP，避免拥塞流控减少了发送速率，如I.4中所描述。

I.6 应用举例：HTTP事务

图I.3显示了一个HTTP事务鉴别的握手时间。从客户的观点，剩余包交换构成了数据传送时间。



图I.3/G.1030 – HTTP协议交换

采用了I.1中的参考连接，显示两个网络段落，带有NAT、防火墙及一个代理服务器作为中间盒子的例子。我们注意到，代理服务器没有将请求目标存储在高速缓冲存储器中，并且对远程Web服务器发出请求。

I.6.1 握手时间

有二种通过不同途径进行的请求—响应，DNS解析时间和TCP三次握手（此处SYN和SYN-ACK是定时的，假设最后的ACK与HTTP GET一起进行，并且不增加明显的时间）。尽管代理服务器分离了TCP连接，这个例子中我们将它作为单个连接上的附加时延来对待。

因为大多数累积是简单的加法和乘法，我们简单地将数值列表并且显示总值（而不是写出带有很多新参数的公式）。

首先，在表I.3中显示DNS性能的成分。

表I.3/G.1030 – DNS解析时间

客户到DNS	RTT (净) [ms]	HPT (处理时间) [ms]	RT (1-p)	每数据流带宽
LAN	2		0.99999	5.0E+6
NAT/防火墙		4	0.99999	
链路1	10		0.9999	5.0E+6
NAT/边缘路由器			0.99999	
NAT到DNS的链路	4		0.99999	10.0E+6
DNS		6	0.99999	
总和 (反转丢包)	16	10	0.00015	

最后，在表I.4中显示了TCP连接建立的成分。

表I.4/G.1030 – TCP连接时间和数据传送时间的客户-服务器路径

客户到服务器	RTT (净) [ms]	HPT (处理时间) [ms]	RT (1-p)	每数据流带宽
LAN	2		0.99999	5.0E+6
NAT/防火墙		4	0.99999	
UNI链路1	10		0.9999	5.0E+6
NAT/边缘路由器			0.99999	
NAT到代理服务器链路	4		0.99999	10.0E+6
代理服务器		4	0.99999	
NNI链路	20		0.99999	10.0E+6
网络段落2	14		0.99999	10.0E+6
UNI链路2	10		0.99999	3.0E+6
防火墙		2	0.99999	
LAN	2		0.99999	5.0E+6
服务器 (TCP处理器)		1	0.99999	
总和 (反转丢包)	62	11	0.00021	

采用DNS解析时间= 16 + 10 = 26 ms, TCP开放时间=62 + 11 = 73 ms, 握手时间99 ms, 并且此时间对大于99.9%的HTTP事务表现出是有效的。

I.6.2 数据传送时间

第一步是确定客户—服务器丢包是否将在此计算中起作用。

再次见表I.4, 我们看到, 最低的每数据流带宽发生在UNI链路2上, 为3 Mbit/s。在12 000比特 (1500字节) MSS, 时延 × 带宽的积为 $0.073 \times 3M = 219$ kbit/s或18.25个包的最佳窗口大小。TCP能被协调来接纳这个路径, 因此就没有与窗口大小相关的限制。

采用瓶颈链路带宽, 我们计算由于探测引起的TCP固有丢包率为:

$$\sqrt{p_{\text{Probing}}} \approx \frac{MSS}{RTT} \frac{C}{0.75 \times BW}, \text{ so } p = \left(\frac{12\,000}{0.073} \frac{0.866}{0.75 \times 3M} \right)^2 = 0.004003$$

由于计算的网路丢包 (0.00021) 大约是 $p_{\text{Probing}}/20$, 我们确定它可以忽略不计, 并且TCP字节流吞吐量将是 $0.75 \times 3 \text{ Mbit/s} = 2.25 \text{ Mbit/s}$ 。

另外一方面，如果估计的网络丢包率与探测丢包率在相同的数量级，则将它们结合起来重新计算带宽将有些悲观：

$$P_{Total} = 1 - \{(1 - P_{Network}) \times (1 - P_{Probing})\}$$

假设一个带有复杂图形的Web网页，对1 Mbit的传送时间为0.444秒。能将服务器处理HTTP GET (0.01秒) 的时间加起来得到总和为0.454秒。

I.6.3 对用户明显的合计时间

合计是握手时间和时间传送时间的总和， $0.099 + 0.454 = 0.553$ 秒。

I.7 小结

本附录描述了评估IP网络中端对端性能的方法。得到的结果可以用于与用户为中心的性能目标相比较，例如ITU-T G.1010建议书，来评估最终用户对不同多媒体应用的总体满意程度。

I.8 附录一的参考文献

- 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

ITU-T 系列建议书

A系列	ITU-T工作的组织
D系列	一般资费原则
E系列	综合网络运行、电话业务、业务运行和人为因素
F系列	非话电信业务
G系列	传输系统和媒质、数字系统和网络
H系列	视听及多媒体系统
I系列	综合业务数字网
J系列	有线网络和电视、声音节目及其它多媒体信号的传输
K系列	干扰的防护
L系列	电缆和外部设备其它组件的结构、安装和保护
M系列	电信管理，包括TMN和网络维护
N系列	维护：国际声音节目和电视传输电路
O系列	测量设备的技术规范
P系列	电话传输质量、电话设施及本地线路网络
Q系列	交换和信令
R系列	电报传输
S系列	电报业务终端设备
T系列	远程信息处理业务的终端设备
U系列	电报交换
V系列	电话网上的数据通信
X系列	数据网、开放系统通信和安全性
Y系列	全球信息基础设施、互联网协议问题和下一代网络
Z系列	用于电信系统的语言和一般软件问题