

国际电信联盟

ITU-T

国际电信联盟
电信标准化部门

E.812

修正 1
(09/2020)

E系列：整体网络运行、电话业务、服务运营和人为因素
电信业务质量：概念、模型、指标和可靠性规划 – 电信业务模型

评估固定和移动宽带网络端到端服务质量的众包方法

修正 1

ITU-T E.812 建议书 (2020) – 修正1

国际操作	
定义	E.100-E.103
有关主管部门的一般规定	E.104-E.119
有关用户的一般规定	E.120-E.139
国际电话业务的操作	E.140-E.159
国际电话业务的编号方案	E.160-E.169
国际选路方案	E.170-E.179
用于国内信令系统的信令音	E.180-E.189
国际电话业务的编号方案	E.190-E.199
水上移动业务和公众陆地移动业务	E.200-E.229
国际电话业务中与计费 and 结算有关的操作规定	
国际电话业务的计费	E.230-E.249
为结算目的对呼叫时长的测量和记录	E.260-E.269
利用国际电话网作非话应用	
概述	E.300-E.319
传真电报	E.320-E.329
有关用户的ISDN规定	E.330-E.349
国际选路方案	E.350-E.399
网络管理	
国际业务统计	E.400-E.404
国际网络管理	E.405-E.419
国际电话业务质量检测	E.420-E.489
业务工程	
话务的测量和记录	E.490-E.505
业务预测	E.506-E.509
确定人工操作的电路数量	E.510-E.519
确定自动和半自动操作的电路数量	E.520-E.539
服务等级	E.540-E.599
定义	E.600-E.649
IP网络的业务工程	E.650-E.699
ISDN业务工程	E.700-E.749
移动网络业务工程	E.750-E.799
电信业务质量：概念、模型、指标和可靠性规划	
与电信业务质量相关的术语和定义	E.800-E.809
电信业务的模型	E.810-E.844
电信业务的业务质量指标和相关概念	E.845-E.859
业务质量指标在电信网络规划设计中的使用	E.860-E.879
设备、网络和业务性能的现场数据收集和评估	E.880-E.899
其他	E.900-E.999
国际操作	
国际电话业务的编号方案	E.1100-E.1199
网络管理	
国际网络管理	E.4100-E.4199

进一步的详细信息，请查阅ITU-T建议书名录。

评估固定和移动宽带网络端到端服务质量的众包方法

修正1

摘要

终端用户设备、用户端设备及其软件已经发展到更快、更强大，已经能够执行数据收集任务。这就使众包方法能够成为现实，这种众包方法在无需修改现有硬件和软件的情况下，可以增加从终端用户那里收集到的技术参数的数量。

监管机构和服务提供商等参与者已通过一种使用众包方法开始评估端到端服务质量。然而，使用众包方法收集的数据进行评估方法可以以多种方式进行部署，不同的方法给出不同的服务质量（QoS）结果。

ITU-T E.812建议书概要介绍了用于评估固定和移动宽带网络端到端服务质量的不同众包方法。

ITU-T E.812建议书修正1 介绍了附录II（众包方法用例）和附录III（固定宽带网络众包实践途径）。

历史版本

版本	建议书	批准时间	研究组	唯一标识码*
1.0	ITU-T E.812	2020-05-29	12	11.1002/1000/14272
1.1	ITU-T E.812 (2020) 修正 1	2020-09-11	12	11.1002/1000/14489

关键词

评估；众包；数据收集；固定宽带；移动网络；服务质量。

* 欲查阅建议书，请在您的网络浏览器地址域键入URL <http://handle.itu.int/>，随后输入建议书的唯一识别码，例如，<http://handle.itu.int/11.1002/1000/11830-en>。

前言

国际电信联盟（ITU）是从事电信和信息通信技术（ICT）领域工作的联合国专门机构。国际电信联盟电信标准化部门（ITU-T）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ITU-T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

本建议书为简要而使用的“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。“须”或“必须”等其他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他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 2021

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	1
2	参考文献	1
3	定义	1
3.1	他处定义的术语	1
3.2	本建议书中定义的术语	2
4	缩略语和首字母缩写词	2
5	惯例	3
6	概述	3
7	众包式数据收集的类型、技术特性和要求	3
7.1	众包式数据收集的类型	3
7.2	发起众包式数据收集的方法	7
7.3	众包式数据收集要求	9
8	设置场景	13
8.1	数据收集设备	13
8.2	测试服务器配置	14
9	给监管机构的其他指导意见	16
附录I	移动业务分层抽样方法的数值算例	17
附录II	众包方法使用案例	20
II.1	可能的使用案例示例	20
II.2	众包解决方案的服务质量指标和计算示例	22
II.3	数据过滤流程规则示例	23
附录III	固定宽带众包的实用方法	26
III.1	CPE数据收集标准	26
III.2	提供终端用户测量工具	27
III.3	向终端用户大规模部署专有硬件探测器	29
参考书目	30

引言

电信服务提供商通常需要评估其网络的服务质量，因为这是一种验证服务是否充分提供给终端用户的方法。从监管机构的角度来看，服务质量的评估可能也很重要，因为质量改进通常是监管日程上的一个关键项目。收集到的有关质量问题的数据可以引导监管机构提高消费者意识，促进网络基础设施进一步改善。电信网络质量管理中最重要的步骤之一是质量监测和评价方法的定义和实施。完成这项任务有许多解决方案可用，所有这些解决方案在准确性、时间和地理粒度以及成本方面都有各自的折中方法。

众包方法是监测和评估固定和移动宽带网络质量的一种可能方法。其基本假设是直接从终端用户的设备（例如移动设备和用户端设备）收集大量终端用户的数据。然而，有许多供应商提供这些解决方案，这些方案之间可能存在显著差异。不同之处包括如何收集数据（是否是由终端用户发起的，或者是否需要干预）或收集什么类型的数据（是否执行下载测试，或者数据是否只是来自设备/设备的常规使用）。

使用相同数据收集方法在同一市场的多个服务提供者的网络上收集众包数据对于服务提供者之间取得一致的数据收集途径和方法大有好处。

与典型的服务质量评估方法（例如，驾车测试和步行测试）相比，众包方法显著增加了数据点的数量。大量的数据可以增强所得结果的可靠性和代表性。此外，这种方法提高了资源的使用效率，并使地理面积较大的国家能够利用公众收集数据。

本建议书确定了采用此类服务质量监视方法时应考虑的优点、挑战和注意事项。

评估固定和移动宽带网络端到端服务质量的众包方法

修正1

编者按：这是一份全文出版物。本修正中的修改内容以与ITU-T E812建议书（2020）有关的修订标记的形式显示。

1 范围

本建议涉及如何使用众包方法评估固定及移动互联网接入的端到端服务质量，包括：

- 固定和移动互联网接入众包方法概述；
- 众包式数据收集的类型、特征和需求；
- 设置场景；
- 在使用众包方法时，可用于基准测试和网络改进的监管机构、服务提供商和供应商指南。

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。所有的建议书和其他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他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 E.800] ITU-T E.800建议书（2008），有关服务质量的术语定义。

[ITU-T E.806] ITU-T E.806 建议书（2019），测量活动、监测系统和抽样方法，以监测移动网络中的服务质量

3 定义

3.1 他处定义的术语

本建议书使用下列他处定义的术语：

3.1.1 下载 (download) [ITU-T E.800]：服务器或主机向个人计算机或设备传送数据或程序。

3.1.2 延迟 (latency) [b-ITU-T G.9961]：从一帧的最后一位通过发送器协议栈的已分配参考点发送的时刻起，到整帧到达接收器协议栈的已分配参考点的时刻止的延迟量度。平均延迟的估计值和最大延迟的估计值都假定为在所有延迟测量值的第99个百分位上进行计算。如果为特定流启用了重传，则延迟还包括重传时间。

3.1.3 抖动 (jitter) [b-ITU-T G.9961]: 一个用于表示等待时间在平均等待时间值之上和之下变动的量。最大抖动定义为在平均等待时间值之上和之下的最大等待时间变动

3.1.4 IP包丢失率 (IPLR) [b-ITU-T Y.1540]: IP包丢失率 (LPLR) 是指观测对象中所丢失数据包数量占所发送数据组的比率。

3.1.5 众包 (crowdsourcing) [b-ITU-T P.912]: 通过一大群人 (最有可能的是在线社区) 获得所需服务。

3.1.6 端到端质量 (end-to-end quality) [ITU-T E.800]: 与通信系统的性能有关的质量, 包括所有终端设备。

3.1.7 服务质量 (quality of service) [ITU-T E.800]: 电信服务满足用户明示和隐含需求的能力的总体特征。

3.1.8 众包式数据收集 (crowdsourced data collection) [ITU-T E.806]: 一种从大量终端用户设备收集主动和/或被动服务质量量度的方法。

3.2 本建议书中定义的术语

无。

4 缩略语和首字母缩写词

本建议书使用下列缩略语和首字母缩写词:

API	应用程序编程接口
CDN	内容发布网络
CPE	客户端设备
CPU	中央处理器
GNSS	全球导航卫星系统
HTTPS	超文本传输协议安全
IoT	物联网
IP	网际协议
ISP	互联网服务提供商
IXP	互联网交换点
KPI	关键性能指标
MCC	移动台国家代码
MNC	移动网络代码
OTT	基于互联网的服务
PoP	入网点
QoS	服务质量
RAM	随机存储器
RF	射频
RoI	投资回报率
UDP	用户数据报协议

5 惯例

无。

6 概述

[b-ITU-T P.912] 在其附件I中将众包定义为“通过一大群人（最有可能的是在线社区）获得所需服务。”从这个意义上说，众包的主要特点是涉及大量的人。这个概念可以应用于各种各样的目的，如筹集资金、分享任务或从意见测验中汇集意见。

当涉及到监控固定和移动宽带网络的服务质量时，众包方法可以理解为从大量终端用户收集数据，目的是评估已使用的网络基础设施的服务质量，以此作为服务消费者需求的一种手段。这可以通过使用直接从终端用户的终端设备（例如，路由器）或移动设备（例如，平板电脑和智能电话）上收集服务质量相关数据的软件和硬件来实现。在固定宽带的情况下，通常由服务提供者提供的用户设备中的嵌入式软件可以收集与服务质量有关的信息。从移动服务的角度来看，也可以收集类似的数据集，例如，通过可在移动设备上下载的软件来收集。

当使用有助于评估固定和移动宽带连接的服务质量的众包解决方案时，可以从用户设备 and 移动设备（本建议书中统一称作“数据收集设备”）上收集大量数据。这些数据包括但不限于：

- 位置（例如，GNSS导航卫星系统坐标），
- 日期和时间（例如，持续时间等），
- 服务提供者（例如，固定宽带移动和互联网服务提供者所使用的家用网络和受访网络），
- 网络信息（例如，接入点名称、信号强度和小区信息），
- 连接类型（例如，以太网、Wi-Fi、3G、4G、5G等），
- 设备信息（例如，制造商、型号等），
- 设备使用情况（例如，用于数据、作为中央处理器（CPU）、电池电量、并发流量等）。

收集到的数据可以为宽带网络提供相关的服务质量指标，如吞吐量、延迟、抖动、数据包丢失等。这些是描述端到端性能的基本参数，可以用来推断服务质量。

7 众包式数据收集的类型、技术特性和要求

众包式数据收集分类为[ITU-T E.806]中给出的主动测量和被动测量。数据收集过程可以由终端用户发起，也可以通过编程方式发起。可以选择不同的配置来满足不同的需求。

基于众包的解决方案可以包括主动和被动测量。混合解决方案可以同时使用两者。

7.1 众包式数据收集的类型

每种类型的数据收集方法都有其自身的优势和挑战，这些将在随后的章节中讨论。虽然两种方法都可以收集服务质量指标，但方法的选择可以有区别地反映同一指标。例如，通过主动测量收集的吞吐量可能更接近网络的性能，而提供被动度量收集的吞吐量可能更接近终端用户的实际使用情况。

7.1.1 主动测量

主动数据收集方法产生人工流量，目的是评估端到端服务参数的质量。例如，的有意的文件传输目的在于测量吞吐量，即ping测试。主动测量方法如图1所示。

主动测量可以是专门为使网络饱和而设计的，因此会产生更精确的度量值，表示测试点的端到端性能峰值。但必须注意对结果进行分析，以确定其可靠性和代表性。为此，第7.3.1和7.3.2节给出了有关数据的恰当处理和统计分析的注意事项。

采用主动测试方法测量端到端性能时，许多因素是可以控制的，例如，测试持续时间、测试服务器位置、并发流的数量、数据包的大小、采样速率或所发送数据的类型（即，随机数据可能不能像文本、图像或视频那样被网络路由器有效地压缩）。

测量上行端到端性能（即，上传）时，主动测试方法可以设计用于测量测试服务器接收到的数据量，而非仅仅是设备发送的数据量。测量数据包丢失率时，主动测试方法可以设计用于对比发送的数据包数量和测试服务器接收的数据包数量。当测量数据包在网络中的通行路径（例如，检测下载路径和上传路径之间的差异）时，只有主动测试才能生成足够的数据包、接收响应并构建使用的路径（例如，使用路由跟踪方法）。

应该意识识别主动测试不能代表端到端性能的场景。在下列场景中，主动测试可能会低估性能：

- 另一进程正在使用数据。例如，如果测试时用户正在发送流视频，主动测试会在其测量的吞吐量范围内受到限制。
- 设备操作系统拒绝网络接入。一般来说，这样做是为了延长电池寿命。在这种情况下，主动测试可能无法发送任何数据，但这并不表示蜂窝网络部分有任何故障。

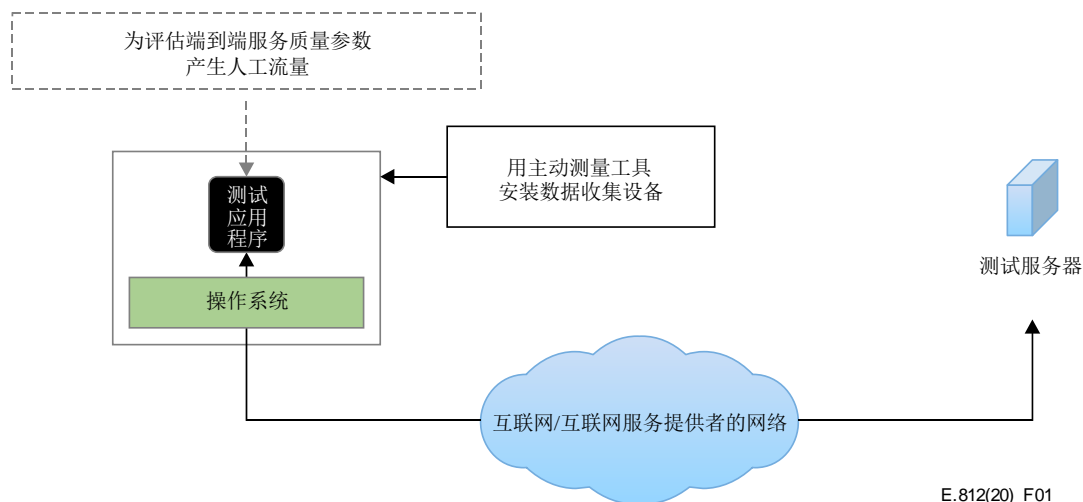


图1 – 移动网络上主动收集数据实例

a) 优点

在主动方法中值得强调的一些优点有：

- 有可能使主动测试标准化。

- 主动测试可以在应用程序层上进行（例如，以测试应用程序、嵌入式工具、网络浏览器的形式运行）。
- 主动测试可以配置为影响不同服务质量参数的不同服务¹的仿真行为（例如，延迟、数据包丢失率）。

b) 挑战

另一方面，还有一些确定的挑战：

- 主动测试使用网络上的资源（例如，尝试测量端到端峰值性能时，向负载很重的网络增加新的负载）和数据收集设备（例如，终端用户的数据配额、电池、随机存储器（RAM）等）。
- 可能会增加高速网络的数据使用量，这取决于测量方法的设计
- 如果设备上的其他进程也在消耗诸如网络连接、CPU和RAM之类的资源，那么用于反映峰值吞吐量的测量会受到抑制。

为了解决上述挑战，可以考虑使用以下技术：

- 对消耗用户流量套餐数据的测试中使用的数据进行零评级可以解决终端用户对数据使用的关注，但使用时应谨慎。使用过程包括向网络运营商提供解决方案特性的细节，如测试服务器或IP地址信息，这可能会导致运营商因集中精力提高测试路径的服务质量参数而使结果不准确。然而，统计工具可以用来分析结果的代表性和可靠性。第7.3.1节讨论了可用的工具，并强调了执行这些分析的注意事项。
- 通过选择测量设计可以使测试不会在更快的网络上消耗更多的数据。例如，模拟特定服务的测试 - 例如流视频（所用带宽为1~1.5 Mbit/s）、游戏或电话业务（带宽通常小于100 Kbit/s） - 可以设计为具有固定的数据消耗量。举例来说，一个测试可能以固定比特率向服务器发送200个包的流，并接收回显数据包。如果正确执行，就可以测量延迟、丢包、抖动和突发速率。然而，无论网速如何，每次测试消耗的数据都是相同的。
- 由于终端用户数量大，则每一个用户需要进行的测量数量较少。解决方法需要在每个设备收集有用的样本数量、避免消耗过多的资源和抑制可实现的用户群规模之间取得平衡。为了实现灵活性，测量抽样策略可以设计为集中配置，这样就可以调整这种平衡。
- 选择峰值吞吐量度量的策略性时机，例如，在终端用户有时间启动数据密集型活动之前，在屏幕解锁后立即运行简短的自动吞吐量测试。
- 采用混合方法（即，主动法和被动法相结合）获得端到端服务质量的整体情况。

¹ 例如：VOIP和在线游戏流量通常涉及以相当稳定和低的带宽传输小的UDP包；而下载电子邮件附件、系统更新和其他文件通常通过TCP进行，这可能会使用更大的数据包和更高的带宽，可能只受到网络的限制。

7.1.2 被动测量

被动数据测量不会通过将人工流量或测试负载注入网络来评估服务质量。相反，这类测量方法更像是无线电参数（终端用户传输的数据）的观察者，并收集关于实际流量的信息。这种方法如图2所示。

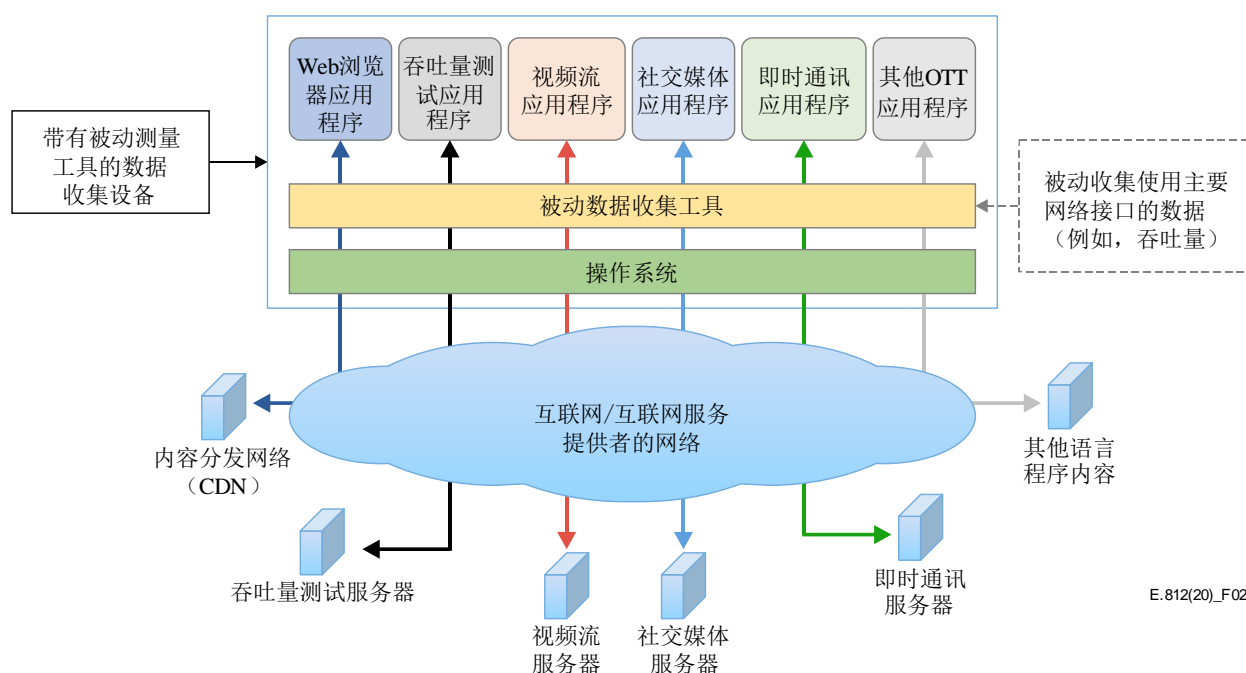


图2 - 移动网络上被动收集数据的实例

被动数据测量可能专为因隐私问题而无法确知产生流量的应用程序而设计，因此，很难准确地判断所做的测量是否是应用程序、网络甚至终端用户的结果。因此，被动数据收集产生的度量值代表用户在该时间点实际使用时体验到的端到端性能。

例如，当被动测量端到端性能时，被动测试无法判断产生流量的应用程序是否以其最大速率产生流量。应用程序可以只在需要的时候发送或接收数据（例如，数据可能被本地缓存），它可以通过计算来延迟一个网络请求，或者用户的操作可能会影响流量（例如，一个用户可能已经暂停了一个视频流）。

同样，在解释被动测试的结果时，传输的数据量、网络状态（例如，移动网络中的信号强度）以及所收集的众包数据集等附加信息可用于更准确地确定用户所体验的端到端性能。

a) 优点

在被动方法中值得强调的一些优点有：

- 执行测量不会消耗额外的数据，这意味着每个终端用户可能会收集更多的样本。
- 人工流量传输不需要维护测试服务器。
- 不会在网络上产生人工流量，因此从流量的角度来看是一种有效的方法。

b) 挑战

另一方面，还有一些确定的挑战：

- 不同的平台（例如，操作系统、硬件等）提供不同的能力为被动测量，因此，难以进行跨平台比较。
- 被动测量不能监测某些服务质量指标（例如，延迟或丢包或终端用户和服务器之间的一致性最大吞吐量）。
- 如果收集应用程序专用数据，则收集数据的方法就具有侵扰性。

为了解决这些挑战，可以考虑使用下列技术：

- 虽然并非所有平台都允许被动测量，但有必要评估排除这些平台是否会显著影响服务质量指标的结果。
- 采用混合方法（即，主动法和被动法相结合）获得端到端服务质量的整体情况。
- 确保终端用户被告知将收集的数据类型，并同意参与众包活动。

7.2 发起众包式数据收集的方法

如果要求终端用户有意开始数据收集，则测试被分类为终端用户发起的。另一方面，如果数据收集是通过预先建立的启动规则以编程方式进行的，则被分类为自动数据收集。

7.2.1 终端用户发起的测量

这种方法要求终端用户发起测试，而测试结果通常提供给终端用户。

a) 优点

这类众包方法具有下列优点：

- 终端用户发起测试意味着用户同意在测试期间进行数据收集，这样就使测试更加透明。
- 终端用户发起测试为其开展更多测试提供选择，并可以在得到许可的情况下在更长的时间内持续进行数据收集，这样就可以收集到更多有关端到端性能和测量状态（例如，设备位置和使用情况）方面的信息。
- 终端用户能够对开展测试的环境进行控制（即，终端用户的位置和网络状态）。
- 终端用户可以对网络的服务质量进行评估，并能在测试后得到测试结果，提高终端用户对测试时段网络状态的认识
- 终端用户发起的测试允许使用调查问题，可以提供关于终端用户互联网连接或测试动机的额外信息。

b) 挑战

依赖终端用户发起测试有可能产生下列挑战：

- 由于这种方法依赖于终端用户的交互，因而收集的样本数量可能会少得多。
- 由于可能有特定的终端用户群体对服务质量评估结果感兴趣，因而可能会产生某些偏差。例如，当终端用户在访问某些应用程序遇到困难时，他们可能倾向于执行数据收集。另一个可能的偏差与收集发生的天数有关，因为没有日程安排。
- 与自动测试相比，恶意用户可能更容易影响所收集的数据，因为终端用户更容易控制测量条件。

- 使用调查问题时，应考虑调查问题的设置，以及问题的设置是可选的还是运行测试所需的条件。测试后问题的设置有可能使答案收到测试结果的影响，而如果在测试前设置问题，则有可能会妨碍测试或得到随意性答案 – 尤其是当问题不能被忽略时更是如此。

7.2.2 自动测量

自动数据收集可以作为一个独立的应用程序或在其他应用程序或硬件中嵌入解决方案，不需要终端用户的干预。测试可以安排为定期运行或基于某些算法或指定规则触发。

a) 优点

这种方法的优点包括：

- 可以确定收集数据的时长和地点。这可能产生大量的样本，有助于实现统计验证。

b) 挑战

自动数据收集可能带来下列问题：

- 自动用户有可能完全意识不到正在进行数据收集。
- 并非所有平台都允许自动化后台度量。

自动发起的测试对设计测试的实体给予更大的控制，这样就有可能确定测试的频次和测试所处的地理区域，这对于完全依赖于终端用户发起的测试的设置来说是无法实现的。

表1对第7.1节和7.2节的情况进行了汇总。

表1 – 众包式数据收集的不同类型和发起方法的优点和挑战

		发起众包式数据收集的方法	
		终端用户发起的众包式数据收集	自动发起的众包式数据收集
众包式数据收集的类型	主动	<p>优点</p> <ul style="list-style-type: none"> - 被认为更加透明 - 提高终端用户的认识 - 有可能使主动测试标准化 - 多平台 - 可配置为模拟服务行为 - 可以设计为在测试期间提供端到端性能的估计 	<p>优点</p> <ul style="list-style-type: none"> - 可以产生大量的样本 - 有可能使主动测试标准化 - 多平台 - 可配置为模拟服务行为 - 可以设计为在测试期间提供端到端性能的估计
		<p>挑战</p> <ul style="list-style-type: none"> - 样本可能很少 - 可能带来偏差 - 使用附加资源 - 有可能增加数据用量 - 结果有可能受数据收集设备状况的影响 	<p>挑战</p> <ul style="list-style-type: none"> - 被认为透明性较差 - 有些情况下不能得到所有平台的支持 - 使用附加资源 - 有可能增加数据用量 - 结果有可能受数据收集设备状况的影响

表1 – 众包式数据收集的不同类型和发起方法的优点和挑战

		发起众包式数据收集的方法	
		终端用户发起的众包式数据收集	自动发起的众包式数据收集
被动	优点	<ul style="list-style-type: none"> – 被认为更加透明 – 提高终端用户的认识 – 可以产生大量样本 – 不需要测试服务器 – 不会进一步阻塞网络 – 根据终端用户的实际使用情况显示端到端性能 	<ul style="list-style-type: none"> – 可以产生大量样本 – 不需要测试服务器 – 不会进一步阻塞网络 – 根据终端用户的实际使用情况显示端到端性能
	挑战	<ul style="list-style-type: none"> – 样本可能较少 – 可能带来偏差 – 有些情况下不能得到所有平台的支持 – 需要监控的服务质量指标有限 – 可能具有侵扰性 	<ul style="list-style-type: none"> – 被认为透明性较差 – 有些情况下不能得到所有平台的支持 – 需要监控的服务质量指标有限 – 可能具有侵扰性

7.3 众包式数据收集要求

本条规定了有关众包解决方案的要求，如认证程序、采样调度方法、数据处理规则等。

7.3.1 采样与调度

根据测试服务器的容量、存储和调度，可以从大量的数据收集设备收集关键性能指标（KPI）数据。

为了进行统计验证，有必要通过考虑预先定义的最大可接受误差来计算最小样本数；前提是总体分布和概率分布相关联（例如，高斯分布用于中位数估计，二项分布用于比例估计）。关于如何获取样本大小以及如何在地理上分配样本的参考文献见ITU-T E.806建议书附件A。

除了最小样本数外，抽样方案必须考虑到有偏差的结果是可以安全地避免的。因此，建议负责收集数据的实体核查样本的地理来源在统计上是否与目标人口的分布情况相一致，是否代表向终端用户提供的实际服务质量。

但是，如有必要，在统计方法中也可考虑地理和/或时间抽样分布。例如，根据活动的目标，将抽样限制在高流量时段，以及在终端用户密度高的地理区域收集更多的样本可能是有意义的。因此，为考虑终端用户基数分布的每个运营商定义抽样计划可能更合适。在其他情况下，可以考虑在流量低的时期或密度低的地区收集更多的样本，以得到比典型的过采样样本组更大的样本容量。

总之，建议确定需要定期和随机监测的终端用户的最低数量。根据时间和地理分布列入数据收集计划的终端用户越多，评价和编制报告就越准确。

然而，由于测试环境不受控制，众包收集的数据必须进行后处理，以丢弃含有噪声的测试结果。例如，出现这类结果的原因有繁忙时间测试计划的影响、终端用户对合同计划的更改、有限制的商业计划、接入技术的变化，等等。在这种情况下，应用丢弃规则后的剩余样本集必须符合统计验证要求（样本数量、估计误差和地理/时间分布）。

在众包式数据收集有一定控制的情况下，对于预先定义的估计误差，可以采用计算样本大小（最小样本数）的抽样方法。例如，在众包测试活动中，如果提前计划，样本的地理分布可以在一定程度上知晓。

然而，在其他情况下，当对地理/时间分布的控制有限时，结果可能不具有代表性，并且可能在估计的KPI中有很高的偏差。例如，当查看整个城市的结果时，可能会发生大多数样本来自单一街区。在这种情况下，结果可能代表了该地区，但不是整个城市。

策划一项测试活动时，必须考虑结果的统计代表性。为此，有定义明确的统计公式可用于推导一组样本的估计误差。[ITU-T E.806]附件A第A.1.2节给出了有关给定样本容量（样本数）置信并考虑观测参数（下载/上传速度、延迟等）的均值和标准差时，用于获得误差估计的样本随机采样一般公式的参考文献。然而，要了解一个地理区域的一组样本的代表性如何，必须计算样本空间分布的估计误差。分层抽样为估计地理代表性提供了必要的公式（见[ITU-T E.806]和[b-Scheaffer]附件A第A1.2节有关分层抽样的公式）。

在分层中，所研究的地理区域（例如城市）必须被划分为不同的层次或地理区域组。每一组都有不同的特征，可以推断出一组地区（通过众包数据样本）对整个地区的代表性。

为了实现这一目标，可以使用归堆技术将一个特定区域划分为更小的区域。堆的大小可以根据研究区域的大小（数十米或数百米或公里）来选择。完成归堆后，就必须把这些区域归入一层。最后，利用分层抽样的公式可以估计出该地区需要具有统计代表性的最小堆数。附录I给出了一个用于说明这种分层方法的数值示例。

通过使用上述技术，我们可以根据所需的样本大小（简单随机抽样）估计最小的堆数（分层抽样），从而确保某一给定的估计误差。

然而，当众包数据来自用户发起的测试或被动数据收集时，不可能提前计划，从而导致没有和/或太少测量值的数据箱（低于95%置信度）和/或测量值高度集中的数据堆。因此，当没有计划/调度可能时，需要统计工具来计算一组样本的估计误差。

在这种情况下，也可以从公式中得到相反的结果；我们可以根据特定城市中带有样本的统计堆数量来计算估算误差。一种方法可以是给数量不足/数量过多的统计堆分配不同的权重，这样结果就可以考虑每个统计堆可能具有的不同数量的测量值。

总之，考虑到众包样本的地理分布所带来的挑战，如果要汇总某个城市/地区的结果，就必须进行统计分析，以了解样本的可靠性，例如，相关的估计误差和可靠性水平。

7.3.2 数据处理

本节概况描述了对可靠分析所需的正确数据处理方法进行了概要描述。

数据集因收集数据的方式而有所不同。分析的目标也会有所不同。由于这些原因，不能禁止任何单一的程序。但是，考察处理步骤的种类和每种步骤的一些示例是有用的。

以下是一些已经确认了的处理步骤：

- 过滤
- 分类
- 聚合

一个典型的处理方法可能包括一组过滤步骤，随后是一组分类步骤，然后是最终的聚合；然而，不同类的步骤也可能交织在一起。

7.3.2.1 过滤

过滤过程可用于删除冗余、不可靠或不相关的数据，这些数据应立即从分析中排除。具体示例包括：

- 重复数据
- 未在待查样本期内收集的数据，或没有有效时间读数的数据
- 未在待查地理边界内收集的数据或没有位置读数的数据

不同的用例情况可能需要不同的过滤器。例如，对于主动测试而言，如果设备的操作系统或设备用户选择权有可能妨碍或限值网络访问，则建议减少某些测试 – 具体的情况如：处于节电模式或飞行模式；或用户手动禁用数据。

7.3.2.1.1 过滤异常测试行为

此处需要特别讨论进一步过滤目标值。如第7.2.1节所述，恶意方可能试图影响结果，例如故意在其知道将产生有利测试结果的时空地点或知道结果较差的其他地点运行测试。通过识别具有不寻常测试模式（如大量的测试，或产生非典型的结果）的设备，这些数据可能会被孤立。

虽然自动化技术在此可以发挥作用，但影响数据的尝试可能会不断变化，以避免具有自适应分析功能的过滤器。特别是，当公众对众包结果进行频繁而细致的公开报道时，用户就有可能判断他们影响结果的尝试是否成功，以及在何种程度上成功。如果用户看到他们的尝试被过滤掉了，他们可能会改变其策略。对于长期或正在进行的有公开报道的众包运动，可以现实地预期，随着应对措施地发现，措施也需要不断开发。

反常的测试行为并非总是被设计用来影响总体结果。用户有可能就打算使用导致产生不寻常数据模式的众包应用程序。举例来说，某些用户可能想提供证据来解除其已有合同，方法是在家里选择数据非常慢的地方运行测试和截屏。另一个极端是，他们可能希望在社交网络上分享自己网速有多快的证据，这导致他们倾向于在网络体验更好的地方进行测试。

通过调查用户在众包代理内部的测试动机（见第7.2.1节），有可能过滤掉此类情况。这类调查应简单、简短、可选，这样有助于提高参与度（这样的定位可以使众多用户将看到并回答：“告诉我们为什么你今天测试，然后开始测试”后，就可以参与调查），避免测试量减少，如果调查是测试的先决条件，则这样的结果是可预期的。即使在大多数用户没有回答这个问题的情况下，如果收集到足够的响应，这些响应也可以作为一个训练集来确定用户的一般动机。

7.3.2.2 分类

分类是对数据需要进行增强或修正，以便为进一步处理做好准备时，需要进行分类。可能有用的分类步骤包括：

- **网络ID映射：** 移动网络有时能够广播不同的网络名称或使用多个移动网络代码（MNC）运行。在这种情况下，将它们映射到规范的标识符和名称可能会有所帮助。如果SIM卡上的移动国家代码（MCC）与所连接网络的MCC所代表的国家不同，则应特别注意，根据正在执行的分析，删除这些国际漫游结果可能是合适的。
- **测量时间校正：** 由于时间可以在终端用户设备中手动调整，设备上报的时间可能不可靠。当设备发送的时间数据与收集服务器的权威值显著不同时，可以应用校正功能。
- **地理编码：** 可以使用纬度和经度形式的位置坐标来分配位置ID，这可以使用层次空间索引，或者一组特定的多边形，例如，表示城市和省份的多边形。

7.3.2.3 聚合

聚合步骤将诸多测试结果置于一个数据集。通常，这些数据将经过过滤、分类，并用于生成最终的汇总统计数据。

所选择的聚合方法在很大程度上取决于分析。应考虑的因素包括：

- 是否应首先将结果聚合到设备用户级别；然后再取这些结果的平均值呢？这种“一个用户一票”策略有可能防止出现数据样本受报告更多数据样本的用户设备过度影响的结果。
- 有必要在任何其他层（地理位置上或时间上）次上对数据进行重新加权？

关于第二点，附录I提供了一个如何在移动服务环境中应用地理分层抽样的例子。

请注意，所有涉及数据收集和数据处理的实体，即供应商、运营商和监管机构，都必须遵守相应的数据保护法规。这与数据处理有关，因为聚合通常需要唯一的用户设备标识符，在某些法律框架中，这些信息与位置数据一起构成个人数据。

8 设置场景

如前所述，众包方法可以分为主动法或被动法，还可以分类为用户发起的或自动发起的数据收集。除众包解决方案法外，根据众包方法实现服务质量数据收集还有其他不同的可选配置。从这个意义上说，必须确定与设置场景有关的两个问题：（i）嵌入众包解决方案的设备，（ii）放置测试服务器的网络点。有必要强调的是，被动众包式数据收集解决方案不需要测试服务器，因为不产生人工流量或测试负载。终端用户发起的或自动发起的测试具有在定义设置场景时必须考虑的专有特征。

这些定义应充分考虑与服务质量评估目标有关的各因素之间的折中。本节中，我们进一步分析了在固定和移动宽带网络中众包式数据收集的不同设置场景的特点

8.1 数据收集设备

众包法数据收集可能发生在不同类型的设备上，这些设备分属服务提供过程的不同元素，[ITU-T E.800]介绍的端到端服务质量模型解释了这一点，如图3所示。

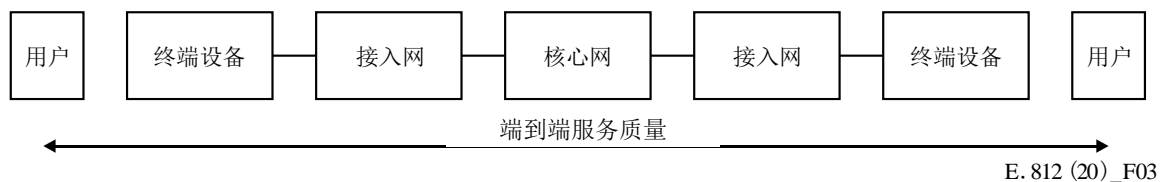


图3 – 端到端服务质量评估模型 [ITU-T E.800]

注1 – [ITU-T E.800]中写明，上图所示配置用于传统业务，用户位于连接的两端。本建议书中针对图3所示的用户使用终端用户这一术语

因此，在评估端到端服务质量时，数据收集解决方案必须分配到终端用户的终端设备中。

对于固定宽带网络，这就意味着将众包解决方案嵌入到用户端设备（CPE）或嵌入到终端用户的设备中（例如，移动设备、个人计算机、智能电视等）。对于移动网络，解决方案是分配到移动设备中的。设备中植入解决方案会带来某些问题，例如对收集结果产生影响的环境变量，比如执行测试位置处的Wi-Fi信号强度。

分配到移动设备中的解决方案可以是专门用于质量评估的应用程序，通常由专门的供应商提供。另一方面，解决方案可以嵌入到任何应用程序中，例如运营商的客户服务程序（前提是运营商有其自己的解决方案）或第三方应用程序，例如社交媒体、游戏和图片编辑器。

分配给个人计算机的解决方案可以是终端用户能够访问的测试网页或web浏览器上运行的后台应用程序或其他应用程序。

特别地，对于固定宽带场景，当众包解决方案分配到终端用户的设备中时，服务质量评估结果与设备有关，这样，设备的技术特性和连接特性（例如，Wi-Fi或以太网）都有局限性。另一方面，如果解决方案直接嵌入到CPE中，则服务质量信息从整体角度反映所提供的服务。

8.2 测试服务器配置

用于承载主动测试的服务器对结果有很大的影响，因此应该谨慎选择。一个好的出发点是考虑服务器托管的类型，还要考虑关于广泛使用的服务器托管的概要总结（见表2）。

表2 – 测试服务器托管选项的比较

服务器托管类型	特点	典型服务
网络接入点 (PoP)	服务器位于服务提供商或连接实体（例如，大学、私人企业或政府机构）的网络基础设施中，意味着接近终端用户。 这里托管的测试服务器用于测试接入网的性能。	托管： <ul style="list-style-type: none">- 文件- 网站
内容发布网络 (CDN)	内容资源通常分布在互联网上的数百或数千个节点上，其设计目的是为内容提供服务。 CDN是一种无服务器的计算形式，可能有若干层，经常访问的内容被缓存在许多边缘节点中。 某些CDN允许创建特定的有应用程序，用于执行测试。	发布： <ul style="list-style-type: none">- 视频和音频点播- 移动应用程序目录- 网站内容（图像、文本、Java描述语言文件，等等）
云计算服务器	计算资源通常分布在互联网上的几十个位置，设计用于支持需要可扩展计算资源的一大类服务。 这种类型的服务器托管还包括边缘计算模型，它在更多的位置引入更小的服务器，通常用于要求有较低延迟的应用程序。	托管： <ul style="list-style-type: none">- 云上应用程序- 数据库- 游戏引擎- 视频OTT呼叫、群组OTT呼叫- IoT服务（边缘计算）- 云游戏（边缘计算）

此外，从终端用户到测试服务器的连接也会影响所测量的特性。测试服务器可以分配到运营商网络（网内）或外部网络（网外）。表3列出了服务器访问类型的比较。

表3 – 服务器访问类型的比较

服务器访问类型	服务器托管类型	特点
网内	<ul style="list-style-type: none"> - 终端用户的服务提供者 - CDN缓存服务器 - 云计算服务的边缘计算缓存服务器 	有助于接入网的质量度量或流行服务的或常用服务的质量度量
网外	<ul style="list-style-type: none"> - 连接的实体（例如，大学、私营企业或政府机构） - 除终端用户服务提供商以外的其他服务提供商 - CDN - 云计算服务 	有助于度量所有网络服务和位置（不论是否流行）的质量

根据终端用户在某一给定网络上所访问的网络和数据服务的概况可得有关应使用的网内服务器和网外服务器的测量值的比例。监管机构可以选择符合其需求的网络数据配置文件。在有多个测试服务器的情况下，可以选择自动或手动选择测试服务器。当自动选择测试服务器时，其中一个选项是选择接近终端用户的服务器，这样可以确保包含更少的网段，从而更准确地测量接入网。另一方面，手动选择服务器允许测量对给定实体的访问。

很明显，对于某个有代表性的服务质量而言，服务器分配会因感兴趣的服务不同而不同。正如上一节所述，主动测试可以设计为专门用于不同的服务，就服务器而言，这可能意味着视频点播测试将从流行视频服务所使用的CDN上传送数据，而多人游戏测试则将与游戏引擎常用的云服务器通信。

将CDN作为测试服务器时，应注意以下两点：

- 经常访问的内容可能比不常访问的内容更接近最终用户。因此，主动测试最初可能下载缓存在几个CDN节点上的文件，但一旦测试运行多次，文件可能分发到更多的节点，并开始更快地下载。通过给任何新的众包解决方案留出热身期，就可以避免这些变化点。
- 动态网际协议（IP）地址分配和超文本传输协议安全（HTTPS）测试在应对移动网络运营商不公平地对待测试流量方面能力更强。举例来说，如果一个流媒体视频的测试使用了网络运营商所知道的静态IP，那么一个限制视频流量的运营商可以仅仅为了这个测试而取消限制，而使用HTTPS和静态IP则会让这变得更加困难。

测量端到端服务质量时，对驻留在某个中立位置（例如，与互联网交换点（LXP）混配）的服务器进行测试可能是不够的，因为大多数为了流量可能不会经过该路径。这种测量端到端服务质量的方法可能会引入移动网络运营商无法直接控制的因素。在上一段的例子中，流行的在线手机游戏所使用的最近服务器可能位于另一个国家，因此国际链接的质量（以及链路覆盖的距离）将有助于衡量服务质量。

鉴于终端用户访问的服务种类繁多，这些服务的端点也有相应的多样性，建议将重点放在少数核心服务上，特别是那些很受欢迎且需要良好端到端性能的服务。对于这些服务，可以专门设计度量方法来获取其端到端服务质量。这涉及到选择与真实流量使用的服务最为匹配的服务器分配方案。

9 给监管机构的其他指导意见

本节为尝试使用基于众包的服务质量测量方法的监管机构提供了其他指导意见。

监管机构可以自行收集数据，要求运营商提供数据或根据本建议书的指导意见使用第三方解决方案。

为了使结果能够解释和可比性，适当的定义和使用共同的方法是必不可少的。

为保证结果的中立性和可比性，监管机构应：

- a) 就所选择的抽样方法、服务质量指标及其计算方法和数据处理规则等方面，定义规则，确保透明度；
- b) 确保数据收集结果不受干扰，例如，通过建立众包工具，将数据发送到中央平台，不需要过滤，最好是实时发送，并监控任何表明试图操纵众包工具使用的测试模式；
- c) 执行审计程序，以检查是否符合要求和已有的测试方法；
- d) 要求数据提供者将方法论的变化（包括但不限于设置、取样和数据收集）通知监管机构进行验证。

在使用由运营商实施的众包解决方案是，为了确保结果的可比性，监管机构可能会决定采用审批或审计流程，其中可能包括对所选设置场景解决方案验证。

验证程序可用于确保运营商之间和/或地区之间结果的可靠比较。在这个意义上，验证的目标是评价数据收集解决方案及其根据技术标准、取样、数据处理和数据收集的测试方法等既定要求，以一定程度的准确性和精密度生成样本的能力。

监管机构可以要求服务供应商或测试供应商提供一份操作手册，其中包括使用的测试方法、测试设置场景、收集的数据或测量指标的技术规范，以及使用的数据处理方法。

如果有些信息无法从监管机构自己收集的数据中获得，或者信息可能过期，监管机构可能会要求运营商定期提供这些信息。

注 – 在固定宽带的情况下，CPE中无法获得的信息可能包括终端用户的一些合同信息（这些参数可以包括最大、最小或保证速度），以及支持执行测量所需的技术标准的设备型号。

附录I

移动业务分层抽样方法的数值算例

(此附录并非本建议书不可缺少的部分。)

本附录中，开发了一个数值示例，用于评估移动服务众包数据的地理代表性，以说明如何遵循[ITU-T E.806]附录A中所述方法的抽样公式。

假设必须基于每个城市的平均下载/上传速度发布结果，这就意味着我们将使用众包数据来获得指标。对于来自移动服务的众包数据，讨论样本的空间估计误差是很重要的。换句话说，就是讨论这些样本对一个地理区域的代表性有多大。由于无法控制样本的地理来源，所以只能估计统计空间代表性。

为了实现这一估计，每个城市首先按照100m×100m的区域大小划分成若干统计子区域（子区域的大小可以有所不同，并且应按照城市的大小来确定）。生成这样的子区域有若干方法，其中包括Geohash法。

为了能够从地理/空间的视角推导出样本的代表性程度，每个子区域可以分层为不同的类别（层）。分层过程中的一个重要方面是每个类别中的元素都具有这些分层共同的特性，且与其他类别的特性不同。

所以，在本例中，分层采样的第一步是选择一个标准，以便构建层，然后按照该标准划分子区域。所选择的标准是人口分布。然而，研究过程中还可以选择其他相关变量，例如，用户密度，设备密度等。

下面是一个使用人口密度标准来定义层的示例：

层数	范围	子区域数
1	人口密度 ≤ 2 000	193 760
2	2 000 < 人口密度 ≤ 4 000	260 766
3	4 000 < 人口密度 ≤ 6 000	213 436
4	人口密度 ≤ 8 000	306 937
		974 8990

划分子区域后，利用每个层的对应的描述性变量（例如，均值、下载速度的标准差和方差等）就能确定估计误差的边界（B）。

最后，在有一定估计误差的情况下，利用以下分层抽样公式（详细描述见[ITU-T E.806]的A.1.1节）就能得到代表研究区域（本例中为城市）的子区域总数。

$$n = \frac{(\sum_{i=1}^L N_i \sigma_i)^2}{N^2 D + \sum_{i=1}^L N_i \sigma_i^2} \quad D = \frac{B^2}{4}$$

其中：

L 是层的总数（用于划分国家的各群；如果划分为城市和农村， $L = 2$ ）

σ_i 是第 i 层的预计标准差；

N_i 是每个层中的地理区域数量（被划分为城市或农村的聚居区数量）；

N 是地理区域的总数（一个国家聚居区的总数）；

$D = \frac{B^2}{4}$ ，其中，其中 B 为估计误差的边界

$$n_i = n \left(\frac{N_i \sigma_i}{\sum_{i=1}^L N_i \sigma_i} \right) \quad i = 1, 2, 3$$

其中

n_i 是第 i 层待测地理区域的数量（对于 $i = 1$ （城市），待测城市聚居区的数量为 n_i ）；

n 是待测地理区域的总数（待测城市和农村聚居区的数量）。

层数	范围	子区域数	均值	标准差	方差
1	人口密度 $\leq 2\,000$	193 760	5.60	0.66	0.43
2	$2\,000 < \text{人口密度} \leq 4\,000$	260 766	5.67	0.43	0.18
3	$4\,000 < \text{人口密度} \leq 6\,000$	213 436	5.23	0.56	0.31
4	人口密度 $\leq 8\,000$	306 937	5.32	0.60	0.36
		974 899			

在计算出获得一个城市统计显著性所必需的子区域总数后，下一步就是得到代表城市内每层的最小子区域数。

举例来说，将估计误差的边界 B 选定为0.1。还可以从之前的结果中得到该变量，或者还可以根据可接受的误差大小将该变量设定为定值。需要记住的一件重要的事情是，公式对估计误差非常敏感，因此小的 B 意味着较大的样本量。还可以通过为 B 设置不同的值来研究多种场景。

设定 $D = 0.0025$ 后，利用分层公式可以得出， $n = 124$ （子区域总数），然后利用[ITU-T E.806]附件A的A.1.1节所给出的第二个分层采样公式分配给每一层（ n_i 为每一层所必需的子区域数）。

$$n_i = n \left(\frac{N_i \sigma_i}{\sum_{i=1}^L N_i \sigma_i} \right) \quad i = 1, 2, 3$$

层数	范围	n_i
1	人口密度 $\leq 2\,000$	29
2	$2\,000 < \text{人口密度} \leq 4\,000$	25
3	$4\,000 < \text{人口密度} \leq 6\,000$	27
4	人口密度 $\leq 8\,000$	42
		124

按照这种方法，每个城市至少可以用 n 个子区域（按照 n_i 分配）表征。下载速度（或者待研究的其他变量）的结果具有的估计误差为 B 。当然，也可以利用该公式得到相对的数值。由于众包数据预先可用，人们可以通过特定城市中带有样本的子区域数量来计算估计误差。

需要特别注意的是，如果每个子区域有最低数量的样品，也可以被认为是有效的。在这种情况下，可以使用简单的随机抽样来得出带有相关估计误差的样本大小。

在规划测试活动时，可以采用三种方法计算每层的样本规模：比例分配、最优分配和混合分配。

a) 比例分配：

按照与单元成比例的分配方式，每一层都具有相同的采样率：

$$f_h = \frac{n_h}{n} = \frac{N_h}{N}$$

换句话说，层越大，该层中选择的样本越多。

因此，如果所有分层都有相同的权重 $w_k = \frac{n}{N}$ ，则这种分配方式将形成自加权采样方案。

这确保了在同时分析几个变量时结果的稳健性。就方差而言，按比例分层的简单随机抽样比简单随机抽样更有效。

b) 最优分配：

最优分配或内曼分配在总人口层面上优化了该变量总值估计器的精度。

内曼分配法假定在分层和高分散性分层中必须选择较大的样本。

令 C 为调查成本，其中：

$$C = \sum_{h=1}^H n_h \times c_h$$

其中 n_h 为分层 h 的大小， c_h 为分层 h 中某个单一的调查成本，

则：

$$n_h = \frac{N_h \times S_h}{\sqrt{c_h}} \cdot \frac{C}{\sum_{h=1}^H N_h \cdot S_h \cdot \sqrt{c_h}}$$

其中 N_h 为分层 h 的人口数量， S_h 为分层的方差。

c) 混合分配：

满足下列公式时，就可以结合前面两种方法的优点，从而保证低成本的健壮性。

$$n = \alpha n_{prop} + (1 - \alpha) n_{opti}$$

通常， $\alpha = 1/2$ 。

附录II

众包方法使用案例

(本附录不构成本建议书的组成部分)

II.1 可能的使用案例示例

在端到端服务质量测量的背景下，有很多可能采用众包式数据收集方法。这包括但不限于表II.1所示的方法。

表II.1 – 使用案例的众包方法范围

范围	众包式数据收集类型	固定/移动网络
网络覆盖	被动	移动网络
性能监测与基准测试	主动与被动	固定网络和移动网络
投诉验证	主动	固定网络和移动网络
承诺许可检查	主动	固定网络和移动网络
网络规划	主动与被动	固定网络和移动网络
网络优化		

通常，如果需要，可以结合其他可用的网络信息层（例如，频带、基站位置，等等）进行上述范围内的部署。以下使用案例是众包式数据收集方法所有可能的确定用途中的一部分，有的也是摘自[b-CrowdWhitepaper]的信息：

II.1.1 移动网络覆盖

以下用例旨在识别室外和室内环境中可用的移动网络覆盖：

a) 室外

- 1) 有了足够的 datapoints（由人群参与决定），可以为每个移动运营商提供精确分辨率的覆盖地图。这可以及时有效地为每个运营商的网络覆盖提供详细展望。
- 2) 如果有代表性的数据，可以使用它进行投诉验证，并识别端到端服务质量有问题的区域。
- 3) 有关地理区域的用户分布指标。由于某一特定区域的众包数据样本来自用户多的运营商的多个 datapoints，因此可以作为用户在网络运营商之间分布情况的指标。

b) 室内

- 1) 室内投诉验证
- 2) 有可能通过信号电平指标定位和识别较差的室内覆盖范围，尽管信号电平问题同样会影响定位精度，或可能完全阻止覆盖（地下车库的较低层）。
- 3) 进入重要的受限建筑，如医院和公共机构。

- 4) 公共区域和商业街区的用户分布指标。由于某一特定区域的众包数据样本来自用户多的运营商的多个数据点，因此可以作为用户在网络运营商之间分布情况的指标。

II.1.2 性能监测和基准测试

- a) 通过KPI监控（主动或被动）或用户交互（例如，用户在服务会话后的反馈），众包数据可以用来评估不同地理级别（直辖市、城市、地区等）的性能趋势。
- b) 利用涉及已发布服务的服务访问和原因方面的详细（时间上的和空间上的）众包数据，可以很好地从根源上识别特定时间和地点使用特定服务版本的终端用户遇到的性能问题，并帮助找到快速修复方法。
- c) 与仅使用众包数据相比，结合用户交互（如调查和其他测量结果来源，如驱动测试）可以更可靠地比较网络运营商或服务提供商的服务质量。

II.1.3 投诉验证

- a) 允许终端用户投诉并做出回应，如网速慢、移动网络覆盖差或没有网络覆盖（特别是在人口少的地区）、无法打电话、通话中断、没有移动互联网（尽管有覆盖）。所有这些都可以通过进一步调查，并有可能通过众包数据的结果来验证
- b) 在确认与投诉相关的问题后，众包数据也可以用于确定投诉的合理性，并提出可行的解决方案。

II.1.4 检查承诺许可

运营商的移动覆盖义务可能包括推出承诺，如在一段时间内覆盖的人口或地区的百分比。这些义务可以通过运营商的许可或频谱拍卖来实现。有了足够的数据点，众包数据就可以用于监控基础设施和频谱方面的网络部署承诺。另一个例子是，固定运营商承诺向最终用户保证一组最小服务质量参数。众包数据可以用来监控这些参数的性能。

II.1.5 网络规划

根据[b-CrowdWhitepaper]和来自网络规划和测试工具供应商的信息，众包数据在网络规划方面的典型使用案例有：

- a) 扩展和/或延申新技术。在这种情况下，众包数据也可以通过观察用户分布如何随时间变化来描述人类活动（就移动而言）。识别数据使用率和/或用户密度高的区域，可用于选择网络中最优部署地点和密度的范围。
- b) 与地形和地理位置数据一起创建和优化传播模型。在这种情况下，众包数据要求有射频（RF）特性方面的度量，例如（但不限于）信号电平、信号质量、频率使用情况或频带。另外，在这一范围使用的众包数据应准确无误，数据收集过程中不存在人工数据，需要具有较高的粒度（有关采样、调度、过滤、分类和聚合等数据收集要求参见第7.3节。
- c) 优化规划过程中使用的流量模型。

d) 根据质量优化（覆盖率、延迟和吞吐量）与投资回报（ROI）的比例，增加盈利率。

II.1.6 网络优化

利用众包数据进行网络优化需要较高的数据精度和粒度。如果这一点能确实保证，则众包数据可以支持以下使用案例：

- a) 定期（通常是每天）跟踪通过受监督网络提供的服务的使用情况和服务质量的趋势（即使是基于网络的监督解决方案无法检测到的微弱信号），以便预测改进情况，如在正确的地点和时间增加容量。
- b) 确定网络问题的根本原因：众包数据可以用于与服务发布、覆盖范围、吞吐量和/或容量相关的地理网络问题检测。
- c) 对特定目标（例如，地理区域、拥有相同型号移动设备的一组用户）专门收集众包数据，以便更快检测和修复特定的问题。
- d) 用定点和按需测试方式（何时何地出现问题）优化和/或替换盲驱动测试（任何地方）和地毯式驱动测试（全部设备）。
- e) 通过室内终端用户密度可视化移动服务的容量使用情况。
- f) 部署后监测网络优化解决方案的效率。

II.2 众包解决方案的服务质量指标和计算示例

表II.2列出了通过众包方法得到的服务质量指标示例及其特征。指标可以根据不同的技术、运营商、地区和时间框架分别汇总。

需要重点强调的是，基于这些指标的结果得出的结论必须考虑解决方案的类型和设置场景。

表II.2 – 服务质量指标示例

指标	描述	服务	观测结果
下载和上传速度	符合约定的速度（下载和上传）	固定业务	只有当设定约定速度信息可用时才能获得该指标。 使用CPE作为设备的解决方案能给出更合适的结果，因为不受Wi-Fi连接限制的影响。
	符合最低速度（下载和上传）	移动业务	只有当设定约定速度信息可用时才能获得该指标。
	典型速度（下载和上传）	固定和移动业务	计算可以使用均值或中值统计，具体取决于数据统计行为。

表II.2 – 服务质量指标示例

指标	描述	服务	观测结果
<u>延迟</u>	<u>一系列延迟测试（UDP包传输中的往返时间）</u>	<u>固定和移动业务</u>	<u>计算可以使用均值或中值统计，具体取决于数据统计行为。</u>
<u>抖动</u>	<u>一系列延迟测试（UDP包传输）中的往返时间变化</u>	<u>固定和移动业务</u>	<u>计算可以使用范围或高、低百分位的数值作为概况统计量，具体取决于数据统计行为。</u>
<u>丢包率</u>	<u>一系列延迟测试（UDP包传输）期间用户数据报（UDP）数据包丢包百分比</u>	<u>固定和移动业务</u>	<u>计算可以使用均值或中值统计，具体取决于数据的统计行为。</u>
<u>移动接入技术（3G，4G，5G，等等）</u>	<u>使用不同移动接入技术观测到的数据点</u>	<u>移动业务</u>	<u>计算可以使用不同接入技术的百分比或比例</u>

注1 – 通过提供有关最差和最佳结果，前表所列不同群组也可以获得第10个百分位数（低）和第90个百分位数（高）指标，也可能适合于补充均值和中值指标。

注2 – 在比较不同地区和运营商的结果时，也可以将符合设定的最小延迟、抖动和丢包率作为补充指标。

注3 – 对于某些聚合指标，例如（但不限于）：丢包率、抖动和速度，获得时间分区等详细信息有助于补充平均值。例如，具有高突发性的低平均丢包率会话可能比具有平均丢包率均匀分布的高平均丢包率会话提供更差的用户体验。

II.3 数据过滤流程规则示例

使用众包方法进行服务质量评估的一个主要方面是验证收集到的数据，并应用预定义的标准接受或拒绝数据记录。

这些标准的具体例子，也称为丢弃规则，见表II.3：

注 – 第7.3.2节为过滤、分类和聚合等数据处理技术提供了指南。

表II.3 – 数据过滤规则示例

确认的情况	描述	观测结果
<u>重复</u>	<u>表示相同时间和设备的记录</u>	
<u>测量失败</u>	<u>错误标志字段指示测试传导过程中存在错误</u>	

表II.3 – 数据过滤规则示例

确认的情况	描述	观测结果
测量条件改变	对于不能立即完成的主动测试，例如下载测试，其中网络类型或接口发生了显著变化（在蜂窝网络和Wi-Fi之间切换，或在不同的蜂窝技术之间切换）。	适用于分析，目的是衡量特定场景的性能，如仅限移动或Wi-Fi。但是，对于某些分析，这些测试将是相关的，可以包括在内。
下载和上传奇偶对等	只显示下载或上传结果的测试	仅适用于每个下载测试都包含一个上传测试对的解决方案。
测试服务器鉴别	测试服务器与接受的服务器不对应	仅适用于主动测试和预先确定测试服务器的用例。例如，如果国际通信不在度量策略的范围内，则必须排除针对位于国外的服务器的测试。此外，如果目标是测量ISP的网络性能，则有必要排除针对ISP网络之外的服务器的测试。
单位时间窗最大测量值超界	测量值的数量是否超过最大值	对于自动测试，还可以设置最大测试数，以避免收集此类数据。或者，只要统计聚合过程防止这些读数对结果产生不当影响，就可以允许这些数据进入管道。 例如，这可以通过在最终计算之前预聚合到时间-设备统计堆来实现。
数据超出位置和期限范围	显示超出所需范围的位置和/或时间期限的测试。	
无效的或不完整的字段值	包含计算指标（如位置、电池电量、设备标识）所需字段的不完整或无效值的测试	数据字段的可接受规则必须是预定义的。
字段结果冲突	不同字段的结果携带冲突信息。	

表II.3 – 数据过滤规则示例

确认的情况	描述	观测结果
电池电量低	从设备在低电池条件下或电池电量使用在可接受的限度下获得的结果	适用于嵌入终端用户设备的解决方案。 对于自动测试，也可以将其设置为不启动测试的条件。
信号电平低	从信号电平低的设备上取得的结果	适用于嵌入到终端用户设备和无线连接（移动或Wi-Fi）的解决方案。 对于自动测试，也可以将其设置为不启动测试的条件。 然而，对于某些分析，这些测试将是相关的，可以包括在内。
并发流量	被设备中并发流量削弱了的结果	适用于主动测试。对于自动测试，也可以将其设置为不启动测试的条件。

此外，在评估结果是否符合合同规定的水平时（如表I.2中的第一个指标），以下筛选标准可帮助计算给定地区/时间内的指标：

- 所述设备的IP与所假设的运营商的IP地址池不对应；
- 用户的约定水平有变化；
- 用户的租约取消；
- 用户的地址有变化；
- 找不到用户的租约信息。

附录III

固定宽带众包的实用方法

(本附录不构成本建议书的组成部分)

监管机构和运营商在固定网络众包中面临的挑战与在移动网络环境中出现的挑战不同。本附录提供了实施固定网络众包解决方案的方法。

III.1 CPE数据收集标准

对于固定网络宽带众包方法，在定义分配给CPE设备的解决方案时，可以考虑三个标准：

- i) 宽带论坛技术报告069 [b-BBF TR-069]，“CPE广域网管理协议”；
- ii) 宽带论坛技术报告143 [b-BBF TR-143]，“赋能网络吞吐量测试与统计监测”；
- iii) 宽带论坛技术报告471 [b-BBF TR-471]，“最大IP层容量度量、相关度量和测量”。

TR-069是终端用户设备和供应商中央控制器之间通信的应用层协议，为用户端设备（CPE）远程管理提供标准化程序。该协议支持自动配置、软件或固件镜像管理、软件模块管理、状态管理等功能。识别机制包括允许根据特定CPE的要求或集体标准（如CPE供应商、设备模型和软件版本）提供CPE。协议中所包含的识别机制使CPE的服务开通或以各具体的CPE要求为基础，或以诸如CPE厂商、模型或软件版本的综合标准为基础。

TR-069的体系结构使客户网关的CPE和客户办公室/家庭网络中的设备都能进行设备管理。

TR-143最初是为运营商设计的一种工具，能够持续监测网络的性能，防止问题发生，并在问题发生时进行诊断。本技术报告提供了CPE数据模型对象的定义，以便使用TR-069中定义的诊断机制对CPE的IP接口进行性能吞吐量测试和数据监测。

在TR-143的范围内，主动监测被描述为“在受控测试中主动发送或接收数据”。从这个意义上说，所进行的测试是以网络层为中心的，与底层接入网无关。此外，主动监测可以支持网络和CPE启动的诊断。TR-471是宽带论坛（BBF）在主动监测和测量范畴的最新发展，改进了TR-143高速互联网接入评估方法。TR-471提供了有关在BBF IP边缘和CPE上部署主动监测的指南，补充的BBF工作增加了必要的数据库模型来配置和控制TR-471测量。

该协议还提供了可选工具，用来管理可选应用或服务中有关CPE的组件，对此需提高安全水平，特别是在涉及支付的情况下。

在众包的背景下对固定宽带网络进行评估，可以采用TR-069、TR-143、TR-471定义的协议进行数据采集，获得服务质量指标。这种方法允许使用现有的CPE作为测试设备。尽管如此，关于众包式数据收集的建议，包括数据采集、测试调度和数据处理，是可以应用的。

这些协议的优点和挑战包括：

- 优点：
 - 测试与底层接入网无关。
 - 服务开通机制可以实现未来的直接扩展，从而提供尚未包含在本版本中的服务和能力。
 - 许多路由器已经支持TR-069标准（全球超过10亿），因此实施是高效和经济的。
 - 虽然TR-069主要用于运营商，但已经有一些供应商拥有TR-069工具集，可以重新用于监管用途。
- 挑战：
 - 虽然TR-069是广泛可用的，但是旧路由器可能不支持它，尤其应重点考虑的是农村地区。
 - 从监管机构的角度来看，基于这些协议实施测量活动可能具有挑战性，因为这需要运营商和供应商的密切合作。

III.2 提供终端用户测量工具

监管机构和运营商都可能向最终用户提供测量工具，例如速度测试。这些工具可以通过web浏览器访问，也可以安装在计算机或移动设备上。移动设备需要通过Wi-Fi连接到CPE，CPE通过固定接入技术（如ADSL/光纤等）接入因特网。这些工具通常取决于用户发起的或自动化的。这种方法的优点是，这些工具易于使用和被最终用户理解，并且可以提供即时的信息和帮助。与移动众包场景不同，测量的可靠性可能会受到客户端设备和CPE之间存在无线链路的影响。众包解决方案通过几种不同的方式来解决这个问题：

- 防止在无线链路上运行测试：这只有在操作系统中安装了速度测试工具并能够检测用户是否通过以太网或无线连接的计算机上才可能。
- 终端用户设备上的测量工具连接到CPE上的专用应用程序编程接口（API），以确定无线链路的质量，并能够通知终端用户测试是否受到了无线链路的影响。CPE上的API还可以提供关于连接到同一CPE的其他设备的带宽利用率的额外信息。此信息对于调整测量结果以补偿其他设备所占用的额外容量或完全排除测量结果是有用的。
- 在计算机或移动设备上本机运行的测量工具能够根据操作系统提供的参数评估设备和CPE之间的本地无线链路，并执行其他测试来验证测量的吞吐量值。其他主动测量使用UDP协议来测量，例如，IP包发送比特率可以估计Wi-Fi容量，并通知终端用户测试结果是否正确地代表了他们的因特网连接速度。

选择如何向用户提供工具将影响在测量过程中可以检索到关于设备和网络状态的数据。表III.1列出了可以根据所提供的工具类型来收集的数据。

表III.1 – 用不同工具收集的数据

	基于web的工具	原生应用程序	CPE上的API
设备信息	受限	是	受限
用户的ISP	是	是	是
固定连接型（例如， 光纤/ADSL）	否	否	是
用户约定的频率方案 （例如100 Mbit、 1Gbit）	否	否	是
路由器型号	否	是	是
CPE已连接的设备列表	否	受限	是
本地背景流量带宽	否	受限	是
Wi-Fi吞吐量	否	是	是
因特网吞吐量	是，但不同的工具可 能产生不同的结果	是，但不同的工具可 能产生不同的结果	是，但不同的工具可 能产生不同的结果
Wi-Fi IP层容量	否	是	是
因特网IP层容量	是	是	是
Wi-Fi延迟	否	是	是
因特网延迟	是，但不同的工具可 能产生不同的结果	是，但不同的工具可 能产生不同的结果	是，但不同的工具可 能产生不同的结果
Wi-Fi连接属性（频 率、信道等）	否	是	是
检测设备限制的可能 性。（例如CPU、 RAM）	否	是	否

根据表III.1可以得出结论，CPE上的API是固定众包的最佳选择，因为它提供了最广泛的收集数据。然而，还必须考虑为所有希望利用该系统的用户部署此类解决方案的复杂性。

从这个角度来看，最合适解决方案是基于web的解决方案，因为该解决方案可以部署在web页面上，所有用户都可以轻松访问该解决方案。

原生应用程序在收集数据量和部署复杂性之间取得了良好的平衡。各种操作系统的原生应用程序确实需要用户克服一个小障碍，但一般来说，用户知道如何在自己的设备上安装新软件。

CPE上部署API复杂性最高，因为它要求分发CPE给用户的路由器制造商或ISP将API集成到每个路由器上。具有最同质CPE部署基础和最先进模型的网络最适合使用API方法。新的CPE模型将有更好的资源来执行新的功能，以及更好的选项来远程升级固件。

表III.1中有一个收集到的数据子集，它对于聚合分析和得出关于服务质量的结论是有用的。此外，还有其他收集到的数据，这些数据对测量本身的验证最有用，在跨不同设备聚合时没有多大价值。

以下是收集到的对聚合分析有用的数据示例：

- 用户的互联网服务提供商（ISP）
- 固定连接类型
- 用户的合约方案
- 路由器模式
- 网络吞吐量/延迟

注 – 在不同的分层上，以不同的方式收集与因特网吞吐量/延迟相关的数据可能会导致不同的结果。因此，必须注意通过混合不同的收集方法来避免不一致的结果。

III.3 向终端用户大规模部署专有硬件探测器

监管机构和运营商可能会决定在客户家中大量部署硬件探测器。通常探测器是一个独立的低成本硬件设备，可以使用可靠的以太网连接到CPE。一个硬件探测器在用户连接的后台运行自动测试，并将结果报告给众包系统的运营商。

在众包场景中，硬件探测的优点和挑战包括：

- 能够安排自动测量，了解网络状况；
- 有可能在专门为运行测量而设计的硬件上可靠地执行不同的测量集；
- 与其他方法相比，硬件探测器由于需要购买特定的硬件和运输而具有较高的成本；
- 由于成本高昂和物流方面的挑战，不能假定所有用户都能利用该系统，这反过来会限制样本量的代表性（第7.3.1条为统计验证的抽样和调度提供指引）；
- 当用户可能停止测量或硬件探测器可能发生故障时，需要持续维护一个可持续的用户群体。

参考书目

- [b-ITU-T G.9961] ITU-T G.9961 (2018) 建议书, 关于统一高速线路的家庭网络收发机-数据链接层规范。
- [b-ITU-T P.912] ITU-T P.912建议书 (2016年), 识别任务的主观视频质量评估方法。
- [b-ITU-T Y.1540] ITU-T Y.1540建议书 (2019年), 网际协议数据通信业务-IP分组传送和可用性性能参数。
- [\[b-CrowdWhitepaper\]](#) [Tobias Hoßfeld, Stefan Wunderer \(2020/03\), 众包网络和QoE测量的白皮书-定义、用例和挑战, Würzburg, Germany, doi: 10.25972/OPUS-20232.](#)
- [\[b-BBF TR-069\]](#) [TR-069的宽带论坛, CPE WAN管理协议 \(CWMP\)。](#)
- [\[b-BBF TR-143\]](#) [宽带论坛TR-143 \(2015/08\), 赋能网络通量性能测试和统计监测。](#)
- [\[b-BBF TR-471\]](#) [宽带论坛 TR-471 \(2020/07\), 最大IP层容量度量、相关度量和测量](#)
- [b-Scheaffer] Scheaffer, Richard & Mendenhall, William & Ott, Lyman (2012). 初级调查取样, 第7版, 圣智学习。

ITU-T 系列建议书

- 系列 A ITU-T 工作的组织
- 系列 D 资费及结算原则和国际电信/ICT 的经济和政策问题
- 系列 E 综合网络运行、电话业务、业务运行和人为因素**
- 系列 F 非话电信业务
- 系列 G 传输系统和媒介、数字系统和网络
- 系列 H 视听及多媒体系统
- 系列 I 综合业务数字网
- 系列 J 有线网络和电视、声音节目及其他多媒体信号的传输
- 系列 K 干扰的防护
- 系列 L 环境与 ICT、气候变化、电子废物、节能；线缆和外部设备的其他组件的建设、安装和保护
- 系列 M 电信管理，包括 TMN 和网络维护
- 系列 N 维护：国际声音节目和电视传输电路
- 系列 O 测量设备的技术规范
- 系列 P 电话传输质量、电话设施及本地线路网络
- 系列 Q 交换和信令，以及相关的测量和测试
- 系列 R 电报传输
- 系列 S 电报业务终端设备
- 系列 T 远程信息处理业务的终端设备
- 系列 U 电报交换
- 系列 V 电话网上的数据通信
- 系列 X 数据网、开放系统通信和安全性
- 系列 Y 全球信息基础设施、互联网的协议问题、下一代网络、物联网和智慧城市
- 系列 Z 用于电信系统的语言和一般软件问题