

骨干网中 VoIP 语音质量的快速评估方法

黄鹂声, 周明天, 汪文勇

(电子科技大学 计算机科学与工程学院, 成都 610054)

摘 要: 提出一种基于被动模式流量分析的 VoIP 语音质量评估方法, 以帮助 VoIP 运营商对主干网中大量并发 VoIP 会话的通话质量进行实时监测。在流量采集的基础上, 通过流跟踪算法获取 VoIP 会话的丢包、时延等基本性能指标; 通过快速评估模型对 VoIP 会话的 R 和 MOS 值进行计算。实验证明, 该方法性能较高, 评估结果准确。

关键词: 被动模式; 流量测量; VoIP; 语音质量; 评估

中图分类号: TP393 文献标志码: A 文章编号: 1001-3695(2008)10-3119-03

Fast method for evaluating VoIP speech quality in network backbone

HUANG Li-sheng, ZHOU Ming-tian, WANG Wen-yong

(School of Computer Science & Engineering, University of Electronic Science & Technology, Chengdu 610054, China)

Abstract: This paper introduced a new passive method for VoIP speech quality evaluation, which helped to monitor service quality for parallel VoIP conversations in network backbone. Based on packet sniffing and flow tracing mechanism, this method measured several performance metrics(such as IPLR, RTT) for each VoIP conversation, and calculated the R and MOS value with a fast model. Test result shows that this method workes with high efficiency and accuracy.

Key words: passive mode; traffic measurement; VoIP; speech quality; evaluation

0 引言

VoIP 已经成为互联网和电信领域的重要应用和服务项目。与传统电话语音相比, VoIP 的语音质量受到 IP 网络服务质量的影响, 存在很大的不确定性, 制约了 VoIP 的发展。因此, 对 VoIP 运营商而言, 如何在全局范围内对大量并发的 VoIP 通话进行质量监测和评估, 是发现问题、改善服务质量的重要基础。

语音通话质量可采用主观方法(如 MOS^[1])和客观方法(如 ITU PSQM^[2]、PESQ^[3]、E-model^[4]等)来进行评估。前者作为一种人为评估方法, 是广泛认同的语音质量标准, 虽然评估结果能够反映大多数人的质量感受, 但成本太高, 费时太长; 后者无须人力成本, 评估结果较为准确, 因此得到更多的研究。在客观评价方法中, PSQM、PESQ 主要基于传统的电话网络, 用于分析个别设备的问题, 而不能反映诸如延时、抖动和丢包等数据网络特有的问题, 没有考虑网络故障对用户感觉造成的影响, 因此不太适合于 VoIP 语音质量评估。目前应用最为广泛的是 E-model 方法, 它是 ITU-T G. 107 提出的一种对电话网络端到端的语音质量进行评估的模型, 关注数据全面的网络损伤因素, 与其他方法相比, 更加适合于数据网络中的语音质量评估。

在 VoIP 的实际应用和运营过程中, 大多数语音通话质量评估方案基于端到端的实现, 即由各个 VoIP 终端、网关设备自行对会话的语音质量进行评估, 评估的结果虽然可以发现端到端的话音质量, 但难以被运营商汇聚和整体把握。针对以上问

题, 本文提出一种基于流量分析的 VoIP 语音质量评估方法, 对骨干网络中的大量并发 VoIP 会话进行质量跟踪。

在 IP 网络中, 影响 VoIP 语音质量的因素包括丢包、时延、抖动等指标。本文首先以 RTP^[5] 为例, 对以上指标的被动模式测量方法进行设计和分析; 在以上度量指标的基础上, 基于 E-model 模型提出一种快速算法, 以完成对语音传输评级参数 R 、平均判分 MOS 指标的实时计算; 最后, 在骨干网络中部署测试环境对以上方法和算法进行验证, 并给出实验结果。

1 基于 E-model 的语音质量评估

E-model 将影响话音质量的若干损耗因素定义为不同的参数, 并将这些综合为一个评级参数 R ($0 < R < 100$), 即

$$R = R_0 - I_s - I_d - I_e + A \tag{1}$$

其中: R_0 表示基本信噪比, 理想状态下 $R_0 = 100$; I_s 包括语音量化和压缩编码所产生的质量损耗; I_d 指时延造成的损耗; I_e 包括由于编码、丢包产生的损耗; A 则表示其他优势因素对质量损耗的弥补。

一般而言, 网络传输过程中不考虑 I_s 和 A 对 R 的影响, 假定其值为 0。因此, R 可以用更简单的方式表示为

$$R = R_0 - I_d - I_e \tag{2}$$

其中: I_d 可以用如下公式表示:

$$I_d = I_{dte} - I_{dle} - I_{dd} \tag{3}$$

I_{dte} 和 I_{dle} 分别表示由于网络时延引起的说话方和收听方回音造成的质量损耗; I_{dd} 则表示由于过长的网络时延带来的质量

收稿日期: 2007-10-26; 修回日期: 2008-03-06

作者简介: 黄鹂声(1975-), 男, 重庆忠县人, 博士研究生, 主要研究方向为下一代互联网络管理及流量工程(lsh@uestc.edu.cn); 周明天(1939-), 男, 教授, 博导, 主要研究方向为计算机网络、网络计算技术、中间件技术、网络与信息系统安全等; 汪文勇, 男, 教授, 硕士, 主要研究方向为计算机网络计算。

损耗。可见, I_d 与网络时延(D) 紧密相关。

影响 R 值的另一个关键因素 I_e 则与编码方式、网络丢包率有关。 I_e 可以表示为

$$I_e = I_p + I_{\text{codec}} \tag{4}$$

其中: I_p 表示因为丢包造成的损伤; I_{codec} 表示语音编码算法造成的损伤。

R 值可以采用以下方式转换为平均判分 MOS^[1]:

$$\text{MOS} = 1 + 0.035R + R(R - 60)(100 - R) \times 7 \times 10^{-6} \tag{5}$$

由此可见, 基于 E-model 的 VoIP 语音质量评估, 只需对每个 VoIP 会话的丢包率、网络时延等基本性能指标进行测量, 并结合会话所使用的语音编码类型, 即可完成对 R 值和 MOS 值的计算。

2 丢包率与时延指标的测量

网络测量方法可分为主动和被动方式。在骨干网络中, 基于流量监测的被动测量方法具有比较明显的优势: 对于 ISP 而言, 只需要一次性部署流量测量探针, 即可对网络中大量端到端会话进行测量, 测量结果易于集中汇聚, 而且测量过程无须终端的参与, 测量成本较低。对网络会话的被动测量原理如图 1 所示。

在被动模式下, 对性能参数的测量方法与主动模式存在较大差别。以下针对部分指标讨论各自的测量方法。

1) 丢包率(IPLR)

方法 1 直接 RTCP 解析。

这是最简单、最直接的方法。VoIP 终端会定期发送 RTCP 包给通话对方, 报告目前的网络质量和统计情况。观测这些 RTCP 包中 SR 报告的丢包率字段, 可以直接获取当前 VoIP 会话的丢包率。该方法虽然简单, 但存在明显的缺点: 丢包率的计算依赖于 VoIP 终端, 而终端的计算可能是不稳定、不可靠的。例如 VoIP 终端一般采用计算收发包数量比率的方法来计算丢包率, 报文的失序和复制数据包会导致计算误差; 另一个问题是 RTCP 包自身的丢失将直接导致测量失败。

方法 2 RTP 序列号连续性观测。

通过对 RTP 包中序列号字段的连续观测可以发现丢包情况。例如在图 1 测量环境中, 连续观测到 i 个 VoIP 终端 A 发送给终端 B 的 RTP 包, 其序列号分别为 $\text{SEQ}_1 \sim \text{SEQ}_i$, 则每观测到一个 RTP 包, 都可以采用如下公式计算出自上一个 RTP 包以来的丢包数量 L_i :

$$L_i = \begin{cases} \text{SEQ}_i - \text{SEQ}_{\text{max}} - 1 & (\text{SEQ}_i > \text{SEQ}_{\text{max}}) \\ 0 & (\text{SEQ}_i \leq \text{SEQ}_{\text{max}}) \end{cases} \tag{6}$$

其中: SEQ_{max} 是到目前为止观测到的最大 RTP 序列号。

任意时刻(观测到第 i 个 RTP 包的时刻) 从某个 VoIP 终端(如图 1 中的终端 A) 到测量点的丢包率可以采用如下方法统计:

$$\text{IPLR}_A = \sum_{k=1}^i L_k / (i + \sum_{k=1}^i L_k) \tag{7}$$

同理可获得终端 B 到测量点的丢包率 IPLR_B 。至此, 估计出终端 A 到 B 的全路径丢包率为

$$\text{IPLR} = 1 - (1 - \text{IOLR}_A) \times (1 - \text{IPLR}_B) \tag{8}$$

该方法计算简单, 但同样容易受到报文失序的干扰。失序的报文会被作为丢包, 导致 IPLR 测量结果偏高。更为严重的

问题是, 丢包率的计算依赖于 RTP 包不断到达的事件驱动, 如果网络路径突然中断, 测量点无法观测到后续 RTP 包, 此时甚至无法对丢包率进行计算。

方法 3 基于预测的统计方法。

该方法的主要思想是: 根据 VoIP 发送端所采用的语音编码类型、语音帧长度, 对 RTP 包的平均发送速率(R_{predict}) 进行预测, 如果单位时间观测到的 RTP 包速率(R_{measure}) 小于预测值, 则它们之间的比值为 VoIP 终端到测量点的丢包率。

$$\text{IPLR}_a = R_{\text{predict}} / R_{\text{measure}} \tag{9}$$

对 R_{predict} 的计算是关键问题。在 RTP 报文中, 可以根据有效载荷类型获取其语音编码方式; 同时, 可以根据连续两个 RTP 报文的 timestamp 字段, 计算出每个 RTP 报文所携带的语音片时长 T , T 的倒数即为该 RTP 会话的 R_{predict} 值。

$$R_{\text{predict}} = 1 / T = 1 / [(TS_i - TS_{i-1}) / R_a] = R_a / (TS_i - TS_{i-1}) \tag{10}$$

其中: TS_i 和 TS_{i-1} 分别为连续两个 RTP 包的 timestamp 值; R_a 表示编码方式对应的采样频率。

相对于前两种方法而言, 该方法不会受到报文失序、RTCP 数据包丢失等情况的影响, 对丢包的统计更加精确。而且, 对丢包的测量灵敏度较高, 即使在网络路径完全中断的情况下, 也能够任意时间窗口内完成丢包率指标测量。

2) 网络时延(RTT)

网络传输时延 D 的值定义为 $T_{\text{start}} - T_{\text{end}}$ 。其中: T_{start} 为数据包传输的起始时间; T_{end} 为包到达目的端的时间。在被动模式下, 由于测量器部署于网络的中间位置, 无法通过发送探测分组的方式来直接获取 T_{end} 和 T_{start} 值, 只能采用间接测量方法。

对于 VoIP 会话普遍采用基于 TCP 的呼叫控制协议, 因此可以通过观测双方的 TCP 报文和相应 ACK 报文的时序关系来对往返时延进行估计。测量原理如图 2 所示。

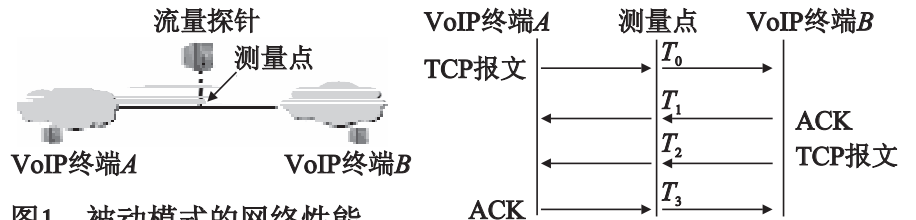


图1 被动模式的网络性能实时测量原理

图2 对往返时延的被动模式测量

A 和 B 两个 VoIP 终端在呼叫和通话控制过程中, 会各自向对方发送信令控制报文, 接收方会尽快返回相应的 ACK 应答。在测量点只需分别对这些报文的到达时间进行记录, 就可以估计双方的路径往返时延:

$$\text{RTT} = (T_1 - T_0) + (T_3 - T_2) \tag{11}$$

其中: $(T_1 - T_0)$ 和 $(T_3 - T_2)$ 分别表示了终端 B 和终端 A 各自到测量点的往返时延。在通话质量出现问题时, 可以很容易地分辨出故障来源。

由于 VoIP 会话的呼叫和控制过程将产生大量交互式 TCP 数据流, 本方法可在 VoIP 存活周期内获取足够多的测量采样, 保证了测量的可持续性和测量结果的稳定性。

3 R 和 MOS 值的计算

在获得 IPLR 和 RTT 指标之后, 理论上可采用式(1) 的方法来计算 R 值, 并采用式(5) 将 R 值转换为 MOS 评分。但式(1) 只是一个概念公式, 不能直接用于 IPLR 和 RTT 到 R 值的

换算。事实上, E-model 本身存在模型化不完全、实例化程度不够的问题, 以致于人们对它能否用于质量测量存在疑虑^[6]。

为解决上述问题, 采用函数拟合^[7]的方式对式(2)中的 I_d 和 I_e 进行实例化, 使其以尽量小的计算复杂度完成对 ITU-T 评估参考值的拟合。

I_d 拟合函数为

$$I_d = 0.024d + 0.11(d - 177.3)H(d - 177.3) \tag{12}$$

其中: d 为网络单向延时, 在以太网环境中, 可以简单地将往返延时除以 2 获得单向延时, 即 $d = \text{RTT}/2$; $H(x)$ 为阶跃函数, 取值为

$$\begin{cases} H(x) = 0 & (x < 0) \\ H(x) = 1 & (x \geq 0) \end{cases} \tag{13}$$

对 I_e 的计算, 则采用与文献[8]类似的对数拟合函数, 将式(4)实例化^[7]为

$$I_e = a \ln(1 + bp) + c \tag{14}$$

其中: p 表示单向传输丢包率, 在前文已经获得; a 、 b 、 c 的值与语音压缩编码方式相关, 其值与部分编码的对应关系如表 1 所示。

至此, 已经获得了丢包率和网络延时指标到 R 值的完整换算过程如下:

$$R = R_0 - I_d - I_e = 100 - 0.024d + 0.11(d - 177.3)H(d - 177.3) + a \ln(1 + bp) + c \tag{15}$$

在测量过程中, 只需要获得每个会话的语音编码方式、IPLR、RTT, 就可以按照上式完成 R 值的计算, 并通过式(5)将其换算为 MOS 分。

4 实验与验证

为了验证上述对 IPLR 和 RTT 指标的测量方法, 本文实现了一个测量原型系统——VPM, 并且选择 CERNET 上的几条主干路径(背景流量在 400 ~ 700 Mbps)进行 VoIP 测量实验。为了验证测量精度, 使用一个改写后的 ping 工具, 通过主动方式进行对比测量。测量过程中, 从 UESTC 校园网向三个不同的远程网络发起 netmeeting 通话。测量结果如表 2 所示。

表1 部分语音编码的Icodec值

标准	a值	b值	c值
AMR(H)	16.68	0.301 1	14.96
AMR(L)	30.86	0.042 6	31.66
G729	21.14	0.127 3	22.45
G723.1	20.06	0.102 4	25.63
iLBC	12.59	0.094 5	20.42

表2 对IPLR、RTT指标的测量结果对比

网络	IPLR值/%		RTT值/ms	
	(VPM)	(ping)	(VPM)	(ping)
1#	0.01	0.01	2	2
2#	1.50	1.38	11.8	13
3#	5.26	4.98	250	275

上述结果表明, 采用本文提出的被动测量方法准确度较高, 与主动测量结果差异不大。

本文以 G723.1(编码速率 6.3 kbps)为例, 根据不同环境下测得的 IPLR 和 RTT 指标对 R 值、MOS 分进行计算, 其结果如图 3、4 所示。

作为对比, 根据 E-model 标准模型^[4], 在不考虑语音编码因素的情况下, 可以根据丢包率和延时指标估计出语音损伤因子, 并由此计算出丢包率、延时指标分别与 MOS 分的关系, 如图 5 所示。

根据图 3 ~ 5 的对比可以看出, 在剔除 G723.1 语音编码造成的固定语音损伤因子之后, 本文提出的快速估计模型的计算结果与标准模型具有很大程度的相似性。

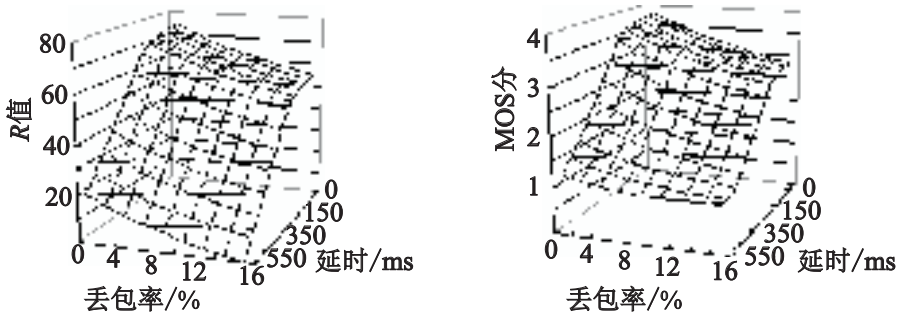


图3 R值与丢包率、延时指标的关系

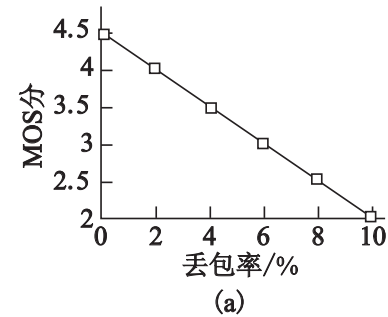


图5 E-Model模型中MOS分与丢包率、延时的关系

5 结束语

本文提出的被动模式测量方法能够对 VoIP 会话的网络质量指标进行实时测量和跟踪, 测量结果能够通过简单模型被换算为标准的 R 和 MOS 分, 从而完成对骨干网络中大量并发 VoIP 会话的实时质量评估。但是, 由于未考虑 VoIP 回音、噪声、端系统缓存等因素, 本文给出的评估结果仅能反映 IP 网络本身对 VoIP 质量的影响。要完成对 VoIP 通话质量的全面评估, 还有许多问题尚待研究。

参考文献:

[1] TU-T P. 800, Methods for subjective determination of transmission quality[S] . Geneva: ITU, 1996.

[2] ITU-T P. 861, Objective quality measurement of telephone-band (300 ~ 3400Hz) speech codecs[S] . Geneva: ITU, 1998.

[3] ITU-T P. 862, Perceptual evaluation of speech quality(PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs [S] . Geneva: ITU, 2001.

[4] ITU-T G. 107, The E-model, a computational model for use in transmission planning[S] . Geneva: ITU, 2000.

[5] SCHULZRINNE H, CASNER S, FREDERICK R, *et al.* RFC 1889, RTP: a transport protocol for real-time applications[S] . [S. l.]: IETF, 1996.

[6] 张薇. VoIP 业务质量评测[J] . 电信网技术, 2005, 6: 80-84.

[7] SUN L. Speech quality prediction for voice over Internet protocol networks[D] . Plymouth: University of Plymouth, 2004.

[8] COLE R G, ROSENBLUTH J. Voice over IP performance monitoring [J] . ACM Computer Communication Review, 2001, 31(2) : 9-24.

[9] DOVROLIS C, RAMANATHAN P, MOORE D. Packet dispersion techniques and a capacity emulation methodology[J] . IEEE /ACM Trans on Networking, 2004, 12(6) : 963-977.

[10] 张冬艳, 胡铭曾, 张宏莉. 基于测量的网络性能评价方法研究[J] . 通信学报, 2006, 27(10) : 74-79.

[11] 谈杰, 李星. 网络测量综述[J] . 计算机应用研究, 2006, 23(2) : 5-8.

[12] 卞静, 王泽强, 张光昭. 基于延迟参数的可用带宽主动估测算法[J] . 电子与信息学报, 2006, 28(10) : 1926- 1929.

[13] 徐加玲, 龚俭. 可扩展高速网络流量被动测量平台的设计与实现[J] . 计算机技术与发展, 2006, 16(9) : 132- 135.

[14] 曹蓟光, 何宝宏. IP 网络性能监测技术及其发展趋势[J] . 金卡工程, 2005, 9(9) : 69-71.