

基于 RTP 协议的 IP 电话 QoS 监测及提高策略

1 引言

随着 Internet 和多媒体技术的飞速发展, Internet 已由早期的单一数据传输网向多媒体数据(视频、音频、文本等)综合传输网发展。但 Internet 提供的只是尽力而为的服务, 不能满足多媒体应用程序对传输延迟、包丢失、抖动控制等要求, 为了能在传统的 IP 网上运行多媒体程序, 必须考虑服务质量(Quality of Service, QoS)。QoS 可用延迟、抖动、吞吐量、丢包率等参数来描述。为了支持网络的实时传输服务, 互联网工作组(Internet Engineering Task Force, IETF)制定了实时传输协议(Real-time Transport Protocol, RTP)。RTP 是专门为交互式音频、视频、仿真数据等实时媒体应用而设计的轻型传输协议, 已广泛应用于各种多媒体传输系统中。IP 电话作为一种新兴业务, 因其低廉的话费受到广大用户的欢迎。但 IP 电话中的通话时延、话音失真一直是制约 IP 电话迅速发展的“瓶颈”。如何确保 IP 电话的 QoS, 是 IP 电话成功与否的关键。

结合 IP 电话系统, 从音频实时传输和控制两方面来讨论 RTP 及实时传输控制协议(Real-time Transport Control Protocol, RTCP)应用技术, 分析影响媒体流实时传输的因素。最后从实际实验、应用的角度, 讨论如何获得当前 Internet 可行的 QoS 监测, 并针对 QoS 质量保证提出切实可行的解决方案。

2 实时传输协议 RTP

RTP 是用于 Internet 上针对多媒体数据流的一种传输协议, 被定义为在一对一或一对多的传输情况下工作, 其目的是提供时间信息和实现流同步。RTP 通常使用用户数据报协议(User Datagram Protocol, UDP)来传送数据, 但 RTP 也可以在传输控制协议(Transmission Control Protocol, TCP)或异步传输模式(Asynchronous Transfer Mode, ATM)等其他协议之上工作。当应用程序开始一个 RTP 会话时将使用 2 个端口: 1 个给 RTP, 1 个给 RTCP。RTP 本身并不能为按顺序传送数据包提供可靠的传送机制, 也不提供流量控制或拥塞控制, 它依靠 RTCP 提供这些服务。通常 RTP 算法并不作为一个独立的网络层来实现, 而是作为应用程序代码的一部分, RTCP 和 RTP 一起提供流量控制和拥塞控制服务。在 RTP 会话期间, 参与者周期性地传送 RTCP 包, RTCP 包中含有已发送的数据包的数量、丢失的数据包的数量等统计资料, 因此, 服务器可以利用这些信息动态地改变传输速率, 甚至改变有效载荷类型。RTP 和 RTCP 配合使用能以有效的反馈和最小的开销使传输效率最佳化, 因而特别适合传送网上的实时数据。

2.1 RTP 数据包

RTP 数据包由 12 个字节的固定 RTP 头和不定长的连续媒体数据(视频帧或音频帧)组成。RTP 协议的数据包格式如图 1 所示。



RTP 报文头部分各个参数的意义如下：

- (1) 版本(V)：2bit 版本号置 2。
- (2) 扩展位(Extension-X)：由使用的 RTP 框架定义。
- (3) 填充(P)：用以说明包尾是否附有非负荷信息。
- (4) 负载类型(PT)：对音频或视频等数据类型予以说明，并说明数据的编码方式。
- (5) 标志位(Marker-M)：标志位由具体的应用框架定义。
- (6) 序列号(Sequence Number)：为了安全，服务器从一个随机初始化值开始，每发送一个 RTP 数据包序列号增加 1。客户端可根据序列号重新排列数据包的顺序，并对丢失、损坏和重复的数据包进行检测。
- (7) 时间戳(Timestamp)：RTP 时间戳为同步不同的媒体流提供采样时间，用于重新建立原始音频或视频的时序。另外，它还可以帮助接收方确定数据到达时间的一致性 or 变化(有时被称为抖动)。
- (8) 同步源标识(SSRG)：帮助接收方利用发送方生成的唯一的数值来区分多个同时的数据流。SSRC 必须是一个严格的随机数。
- (9) 作用标识(CSRC)：网络中使用混合器时，混合器会在 RTP 报文头部之后插入新的同步源标识，其作用是区分多个同时的数据流。

2.2 RTP 控制协议——RTCP

在 RTP 会话中，RTCP 周期性地给所有参与者发送控制包，应用程序或第三方监控者接受 RTCP 控制包，从中获取控制信息，估计当前 QoS，以便进行传输控制、拥塞处理、错误诊断等。

RTCP 报文头部参数首先要区别携带不同控制信息的 RTCP 报文的类型，RTCP 报文的类型主要有以下几种：(1) SR：发送报告，当前活动发送者 发送、接收统计；(2) RR：接收报告，非活动发送者接收统计；(3) SDES：源描述项，包括

CNAME; (4) BYB: 表示结束; (5) APP: 应用 特定函数。其中最主要的 RTCP 报文是 SR 和 RR。通常 SR 报文占总 RTCP 包数量的 25%, RR 报文占 75%。

通过这 5 种控制包, RTCP 协议实现了以下 4 个主要功能:

(1) 提供数据发布的质量反馈, 这是 RTCP 最主要的功能。作为 RTP 传输协议的一部分, 与其他传输协议的流和阻塞控制有关。反馈对自适应编码控制直接起作用。反馈功能由 RTCP 发送者和接收者报告执行。

(2) 送带有称作规范名字 (CNAME) 的 RTP 源持久传输层标识。如发现冲突, 或程序重新启动, SSRC 标识可改变, 接收者需要 CNAME 跟踪参加者。接收者也需要 CNAME 与相关 RTP 连接中给定的几个数据流联系。

(3) 根据参与 RTP 会话的数量调整 RTCP 的发送速率。

(4) 传送最小连接控制信息, 如参加者辨识。最可能用在“松散控制”连接, 那里参加者自由进入或离开, 没有成员控制或参数协调, RTCP 充当通往所有参加者的方便通道, 但不必支持应用的所有控制通信要求。

3 由 RTP 包分析影响多媒体数据流实时传输的因素

随着 VoIP 领域不断发展, 满足网络 QoS 检测需求的应用也成为引人注目的焦点。IP QoS 是指 IP 的服务质量, 也是指 IP 数据流通过网络时的性能。目的就是向用户提供端到端的服务质量保证。有一套度量指标, 包括业务可用性、延迟、可变延迟、吞吐量和丢包率等, 现就项目中在上海阿尔卡特网络支援系统有限公司 NGN 实验室中所得到的 RTP 和 RTCP 包进行分析, 主要研究其中 3 个因素, 从而达到对实时流媒体数据进行监控的目的。

3.1 抖动

抖动会引起端到端时延的增加, 引起语音质量的降低。在音频数据的传输过程中, 由于传输延迟的不稳定而造成相邻数据包接收时刻间隔不稳定, 从而产生抖动。消除抖动的主要依据就是 RTP 包的首部中包含的时间戳字段。时间戳标志着该段音频数据中第一个采样点的采样时间。每两个 RTP 包的抖动可以用其 RTP 包中的 RTP 时戳和接收的时刻进行计算。

关于包的传送时间, 接收者最先了解到的是它的时间戳和接收者当前时间之间的差值。该差值是: $D_i = R_i - S_i$, 表示从包被盖上戳开始, 到它在信源的 输出链路上被实际发送为止, 其中的传送时间和某个机器时间。RFC 1889 建议使用 NTP 来完成端点的端到端同步, 但是也有非同步端点实现存在。

包 i 和包 j 之间增加的延迟差 (二阶效应) 计算公式如下: 设 R_j 代表第 j 个包的接受时刻, S_j 代表第 j 个包的 RTP 时戳值, 则第 i 个 RTP 报文与第 j 个 RTP 报文间的抖动为 $D(i, j)$

$$D(i, j) = (R_j - R_i) - (S_j - S_i) = (R_j - S_j) - (R_i - S_i) \quad (1)$$

在生成 RTCP 报文时，其应当传送的时延抖动的值可用以下公式进行递推计算

$$J(i)=J(i-1)+[|D(i-1,i)|-J(i-1)]/16 \quad (2)$$

其中：J 为要传送的时延抖动值。对后一项除以 16 是为了消除连带噪声。

抖动是分组交换的必然结果，影响抖动的因素一般和网络的拥塞程度有关。由于语音同数据在同一条物理线上传输，语音数据通常会由于数据报文占用了物理线路而导致阻塞。解决抖动通常采用缓冲队列来解决(在网关、IAD 上均有 JitterBuffer 来消除抖动)，每收到一个数据包，先将其放入缓冲区，应用程序在缓冲区另一端取数据，只要缓冲区足够大，抖动一定能被平滑掉。而错序是由于网络拥挤而使某些后发的数据包先到达收端而引起的，只要设置足够大的缓冲区来对数据包重新排序，就能解决这个问题。或者需要 IP 承载网采用 QoS 策略，保证语音数据的最高优先级，得到最先发送和获得高带宽也是解决抖动问题的主要手段。

3.2 时延

时延是处理和传输导致数据不能按时到达的延迟，是影响流媒体数据传输的一个主要因素。话音信号在端到端传输过程中受到的时延迟滞通常包括：编解码器引入的时延、打包时延、去抖动时延、承载网上的传输节点中排队、服务处理时延。这些时延累计的总和将影响话质，导致回声干扰和交互性的劣化。对于 VoIP 系统，规定时延一般控制在 150 ms 内。

分组语音网络中的延迟可分为固定延迟和可变延迟。前者相对容易得到，笔者不作考虑。在计算丢包率时，主要考虑可变延迟。丢包判定等待时限 T_{wait} 设定的大小在很大程度上影响丢包率计算的准确性，也就是可变延迟的影响，它与语音包的传输延迟 T_{trf} 有关， T_{wait} 越大等待时限就越长。但不能超过保证语音流连续播放的时间上限 T_{max} (T_{max} 一般取 250 ms)，即： $T_{wait}=\min(T_{wait}, T_{max})$ 。 T_{trf} 可根据 RTCP 协议的 SR 控制包中的 NTP(Network Time Protocol)时间戳计算得到，见图 2。



使用 RTCP 包计算网络时延要求在两个端点之间传送发送方报告 (RTCP SR)。RTCP 包中的报文类型以二进制表示，其中十进制的 200 代表此包的报文类型为 SR(发送报文)。收到 200 的包类型，表明可以提供足够的信息，计算与这个 SSRC 有关的发送方的延迟。那么就计算发送方流 (SSRC) 和 源流 (SSRC_n) 之间的双向延迟。每次在收到 RTCP SR 包时，必须保存上一个时间戳 (Last SR Timestamp, LSR) 和

DLSR(Delay SinceLast SR)值。SSRC 的 LSR 域是从该 SSRC 接收到的 NTP 时间戳的中间 32 bit, 只要接收到一个 NTP 时间戳就可以了。如果没有接收到 NTP 时间戳, 那么这个域应该设置为 0。信源可能无法获得时钟, 或者其他可接受的消逝的时间。因此, 在会话持续期间, 特定 SSRC 的这个域可能保持为 0。DLSR 域是端点从发送者那里接收到一个 SR 以后过去的时间。这个计数器的每一步代表 $1/65\,536\text{ s}$ 。一般来说, 在 NTP 时间戳, LSR 或 DLSR 等于零时, 不应该计算双向延迟。网络传输时延 Ttrf 即为双向延迟的一半。具体计算方法如下: 设目的网关接收源网关发送的控制包 SR_i, 经过一段时延 DLSR 后, 向源网关发送相应的响应 SR 控制包 SR_j, 并从控制包 SR_j 中的 NTP 时间戳域中提取中间的 32 bit 作为控制包 SR_i 的 LSR。如果没有收到源网关发出的 SR 控制包, 则 LSR 和 DLSR 都置为 0。最后, 将 DLSR 和 LSR 的值填入到控制包 SR_j 的相应域中。设源网关接收到控制包 SR_j 时刻为 A, 则当前网络传输延时为

$$T_{trf} = (A - LSR - DLSR) / 2 \quad (3)$$

由实验包中的数据得出的 Ttrf 是使用 32 bit 归一化 NTP 时间戳计算的, 所以要把延迟转换成毫秒单位, 结果中前 16 bit 将与延迟秒数对应, 后 16 bit 与 $1/65\,536\text{ s}$ 的数字对应。因此, 上面 16 bit 必须乘以 1 000, 把秒转换成毫秒。下面 16 bit 必须除以 65 536, 然后再乘以 1 000。两个因数之和可以得到最终结果。

实时语音传输要求端对端时延不能太大, 端到端的时延要求分成 4 个等级: $\leq 150\text{ms}$ (最好)、 $150 \sim 250\text{ms}$ (较好)、 $250 \sim 450\text{ms}$ (一般) 和 450ms 以上 (差)。可以通过设定 IP 优先级、路由选择、RED (随机早期检查) 等技术来缩短 IP 网络时延。

① 优先级是指对每个数据包的级别进行分类, 不同级别的数据包在网络进行预留带宽分配、通过顺序、时延抖动、丢包等方面处理时, 所受到的待遇不同, 这样可以确保语音、图像等对实时性要求比较高的数据包优先传输, 以提高传输质量。

② 选择合适的路由绕过那些负载过重的路由器, 直接连到主干网进行传输。

③ 当网络拥挤发生拥塞时, RED 就优先丢弃一些对话音影响较小的数据包, 并让终点站降低传输速率, 避免路由器或交换设备缓冲区溢出。

此外, 采用流量控制、队列管理、数据包保护技术也可以提高网络的管理性能, 从而缩小网络时延。

3.3 丢包率

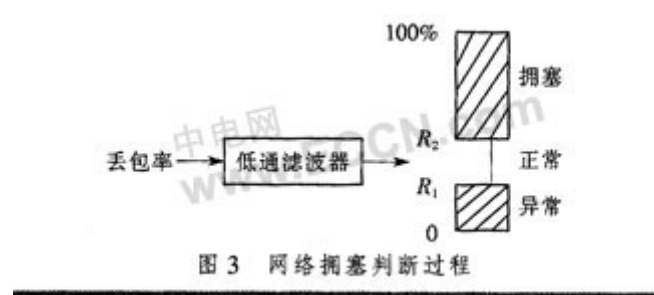
丢包率定义为在网络中传输数据包时丢弃数据包的最高比率。数据包丢失一般是由网络拥塞引起的。丢包对 VoIP 语音质量的影响较大, 在本实验环境中, 采用 G. 711 时, 当丢包率大于 10% 时, 已不能接受, 而在丢包率为 5% 时, 基本还可以接受, 因此, 要求 IP 承载网的丢包率小于 5%。

丢包率为单位时间内丢失的语音包的数目。检测的具体方法是统计语音流中的连续 2 个 SR(发送端报告控制包)的时间间隔 T_{int} 内实际接收语音包的数目 N_{real} ，然后按下面公式来计算丢包率 R_{lost}

$$R_{lost} = (N_{exp} - N_{real}) / N_{exp} \quad (4)$$

其中： N_{exp} 为期望的语音包数目，它等于在 T_{int} 内接收到语音包的最小序号 N_i 和最大序号 N_j 之差，用下面公式来表示： $N_{exp} = N_j - N_i$ ，实际接收到的语音包 N_{real} 可能包括在 T_{int} 之前的迟到包(由于语音数据采用 UDP 协议传输，不存在重复包)，所以计算所得的丢包率会出现负数。此时丢包率定义为 0。

丢包率的计算可在每个丢包判定等待时限内计算，也可按照一定的时间间隔来计算一次，而这个时间间隔要根据网络的负载、网络的稳定情况来确定。为了减少网络的负载，不是每次计算的丢包率都要反馈给源网关的，而是事先根据网络用户期望的语音质量来确定两个丢包率状态的阈值 R_1 和 R_2 。对于音频而言，取 $R_1=3\%$ ， $R_2=8\%$ 。当 $R_{lost} > R_2$ 时，将丢包率信息的标志位，置为“overload”(拥塞)，表示网络处于拥塞状态；当 R_{lost}



根据 RTP 头中的 sequence number 域，可以在接收端很轻易地发现包丢失，为丢包修复奠定基础。在实际使用中发现，绝大多数丢包是单个丢包，两个或两个以上包丢失的比例较小。针对单个包的丢失，传统的丢包处理方法有两种：一种方法是重发，但在传输语音数据时，重发将引起播放质量下降，出现无法识别的话音或回音现象；另一种方法是忽略，这同样会影响播放质量。更好的方案是使用拆分法优化丢包损失。拆分法的基本思想是：在发送端把原来要打入一个 RTP 包的话音数据按照采样间隔分成两块，然后采用相同的压缩算法分别压缩、打入 RTP 包，并标记相同的时印进行传输。在接收方执行相反的过程，把解压缩后的数据采样、合并、回放。

如果某个 RTP 包在传输过程中丢失，那么丢失的只是原数据包按采样间隔的一半信息，接收端可以用接受到的另一半信息，利用插值等方法恢复出原语音包的大部分信息，从而使话音质量不至于下降太多。拆分法的主要思想如图 4 所示。



4 结束语

对音频数据的实时传输问题进行了详细分析，在分析 RTP 协议的基础上，探讨了基于 RTP 协议的 QoS 动态监测的一些方法，并提出了解决在流媒体中存在的语音实时传输质量保证的策略。避免语音通信实时性差的缺点，减小了网络延时使抖动的影响减低，改善了语音传输效果。目前，IP 电话用户数每年正以 239% 的速度增长。下一步将以此为依据设计出基于 RTP 的一个应用模型，进行深层开发研究。