

第一章 RTP 协议

§ 3.1 RTP 协议概述

RTP 协议是由 IETF 开发的, 1996 年成为 RFC 正式文档, 是用于 VoIP, 视频等实时媒体传送的协议。

RTP 为交互式音频、视频等具有实时特性的数据提供端到端的传送服务。在 IP 网络上, 一般是在 UDP 之上运行 RTP 协议。如果支持它的网络能提供组播功能, 则 RTP 也可用组播将数据送给多个目的用户。

RTP 包括两个关系十分密切的子协议:

实时传输协议 (RTP): 用于传输实时数据。

实时控制协议 (RTCP): 用于监视网络的服务质量, 并传递与会者会话中的信息。

下面首先给出有关 RTP 的一些定义。

- RTP 会话 (RTP session): RTP 传输服务使用者之间的连接被称为 RTP 会话, 就每一个会话参加者而言, 会话由一对传输层地址 (即一个网络层地址加上两个端口地址, 一个端口为 RTP 报文的发送/接收所占用, 另一个端口为 RTCP 报文的发送/接收所占用) 标识。在 IP 多播方式中, 每个参与者的目的地运输层地址对可以都相同; 在单播方式中, 每个参与者的地址对均不相同, 因为每个人的网络层地址都不相同。在多媒体会话中, 每个媒体信号由不同的 RTP 会话传送, 有其自己的 RTCP 分组。各 RTP 会话由不同的端口对和/或不同的多播地址区分。

- RTP 媒体类型: 由一个 RTP 会话传送的所有净荷类型的集合。RTP 应用文档指定 RTP 媒体类型和 RTP 净荷类型的关系。

- RTP 净荷: 由 RTP 传送的数据, 如音频抽样信号, 压缩视频数据等。净荷格式及其解释由应用层规定。

- RTP 分组: 由 RTP 头部和净荷数据组成。通常下层传送协议的一个数据包 (如一个 UDP 包) 只含一个 RTP 分组; 如果采用一定的封装方法, 也可以包含多个 RTP 分组。

- 同步源 SSRC (Synchronization source): RTP 包的信源流, 在 RTP 头中用 32 比特长的同步源标识符表示, 与网络地址无关。该字段用以标识信号的同步源, 其值应随机选择, 以保证 RTP 会话中任意两个同步源的 SSRC 标识都不相同。

- 提供源 CSRC (Contributing source): 分信源标识, 32 比特。

RTP 分组头部 最多可包含 15CSRC 标识, 其数目由 CC 字段指明。如果分信源的个数大于 15, 只能标识其中 5 个源。CSRC 标识由混合器插入, 其值就是组成复合信号的各个分信号的 SSRC 标识, 用以标识各个组成成分信号的信源。

- 混合器 (Mixer) 和翻译器 (Translator): 这是 RTP 在接收方和发送方之间引入的两类功能模块。混合器接收来自一个或多个发送方的 RTP 组块, 并把它们组合成一个新的 RTP 分组继续转发。这种组合组块将有一个新的 SSRC ID。因为这些来自不同信源的组块可以非同步到达, 所以混合器改变了该数据流的临

时结构。与混合器相反，转换器只改变组块内容而并不把数据流组合在一起。

混合器和翻译器保证具有不同通信条件的用户可以顺畅地完成数据交换。混合器的重要应用是在高带宽网络与低带宽网络之间，改变从高带宽网络接收数据的压缩编码，把音、视频码流转换成低码流数据发送到低带宽网络；翻译器用在不同协议的网络之间和在防火墙之间，因为 UDP 通过防火墙会发生阻塞现象。

§ 3.2 RTP（实时传输协议）

RTP 协议（Real Time Protocol）提供具有实时特征的、端到端的数据传送服务，可用来传送声音和运动图像数据。在这项数据传送服务中包含了装载数据的标识符、序列计数、时戳和传送监视。通常 RTP 的协议元是用 UDP 协议元来装载的，并利用 UDP 的复用和校验和来实现 RTP 的复用。

必须注意的是，RTP 没有提供任何确保按时传送数据的机制，也没有提供任何质量保证的机制，因而要实现服务质量必须由下层网络来提供保证。同样必须注意的是，RTP 不保证数据包按序号传送，即使在下层网络能使保证可靠传送的条件下，也不保证数据包按序号传送。包含在 RTP 中的序号可供接收方用于重构数据包序列，也可用于包的定位。

在 RTP 会话中，复用是由目的传送地址（网络地址和端口号）来提供的，一个传送地址定义了一个 RTP 会话。例如，在一个会议中，音频码流和视频码流是分别用不同的 RTP 会话（一般是同一个目的网络地址和不同的端口号）来传送，而不是用 PT（载荷类别）或 SSRC 作区分来实现单个 RTP 会话中的复用传送。

固定报头的 RTP 报文结构如图 3-1 所示：

RTP 报头中各字段的含义如下：

版本（V）：2 个比特，表示 RTP 的版本号。

填充（P）：1 个比特，置“1”表示用户数据最后加有填充位，用户数据中最后一个字节是填充位计数，它表示一共加了多少个填充位。在两种情况下可能需要填充，一是某些加密算法要求数据块大小固定；二是在一个低层协议数据包中装载多个 RTP 分组。

扩展（X）：1 个比特，置“1”表示 RTP 报头后紧随一个扩展报头。

CSRC 计数（CC）：4 个比特，表示在定长的 RTP 报头后的 CSRC 标识符的数量。

标记（M）：1 个比特，其具体解释由应用文档来定义。例如，对于视频流，它表示一帧的结束，而对于音频，则表示一次谈话的开始。

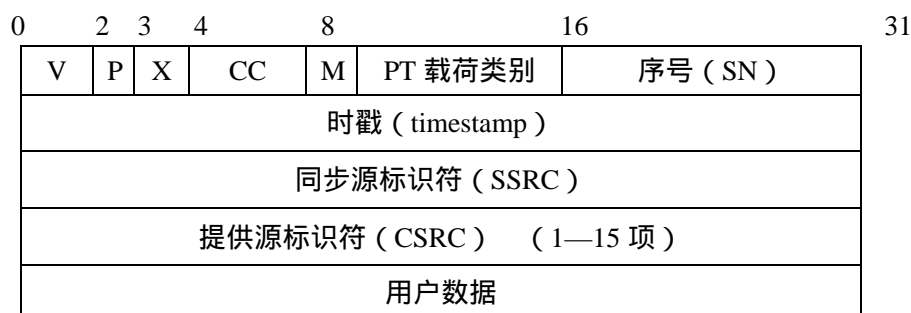


图 3-1 固定报头的 RTP 报文

载荷类别 (PT): 7 个比特, 它指示在用户数据字段中承载数据的载荷类别。

序号 (SN): 2 个字节, 每发送一个 RTP 数据包该序号增加 1。该序号在接收方可用来发现丢失的数据包和对到来的数据包进行排序。

时间戳 (TS): 4 个字节, 它用来表示 RTP 包中用户数据段的第一个字节的采样时刻。时间戳的时间表示应为线性单调递增的, 以便完成同步实现和抖动的计算。

同步源标识符 (SSRC): 4 个字节, 用来标识一个同步源。此标识符是随机选择的, 但要保证同一 RTP 会话中的任意两个 SSRC 各不相同, RTP 必须检测并解决冲突。

提供源标识符 (CSRC): 它可有 0~15 项标识符, 每一项长度为 32 比特, 其项数由 CC 字段来确定。如果提供源多于 15 个, 则只有 15 个被标识。

为了能满足各种应用的需要, RTP 报头可进一步扩充, 其时 X 比特将置“1”, 扩充的 RTP 报头部分则紧随在 SCRC 清单之内。

如果扩展位被置为 1, 意味着 RTP 固定头后紧跟着一个头扩展, 其格式如图 3-2 所示。前 16 位的内容由轮廓文件决定, 主要用来标识不同的头扩展类型。这种扩展方式主要用来传递独立于具体格式的载荷 (payload-format-independent) 的应用信息。

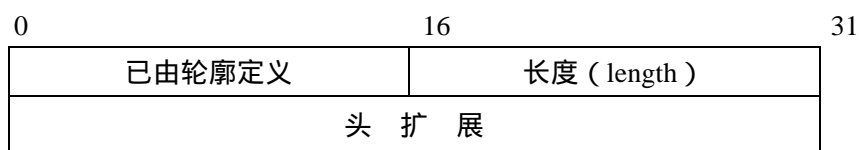


图 3-2 RTP 头扩展

§ 3.3 RTCP (实时传输控制协议)

3.3.1 RTCP 的功能

RTCP 是 RTP 的控制协议, 它用于监视网络的服务质量和在正在进行的与会者会话中传递信息。RTCP 的基本做法是周期性地向会话的所有参加者进行通信, 采用和数据包分配传送的相同机制来发送控制包。RTP 有四个功能:

(1) RTCP 的基本功能是提供关于数据传输质量的反馈, 这是 RTP 协议的重要组成部分, 它大致和其它传输协议的流量控制及拥塞控制相对应。RTP 向会话的所有参加者发送反馈报文, 其目的是允许那些发现数据传输过程中出现问题的用户进一步判断问题仅和其自身相关还是和全体参加者有关, 并采取相应的措施。此外, 在 IP 组播机制的支持下, RTCP 的这一功能允许监视者并不具体参加某一 RTP 会话的功能实体, 而是通过接受反馈报文来推断网络的服务质量, 扮演第三方监测者 (third-party monitor) 的角色。该功能由 RTCP 的发送者报告 (SR) 和接收者报告 (RR) 来具体完成。

(2) 就每个 RTP 源而言, RTCP 为其分配一个永久的传输层标识, 该标识被称为 CNAME。虽然 SSRC 标识符可以区分一个会话中的不同码流, 当 RTP 发现不同码流的 SSRC 标识发生冲突或应用重新启动时, SSRC 值会被更改, 因而接收者需要利用 CNAME 而非 SSRC 来跟踪每一个会话参加者; 另外, 借助于

CNAME,接收者可以关联来自于同一用户、经由不同 RTP 会话传来的码流以完成特定的任务,如音频流和视频流的同步。CNAME 由 RTCP 的 SDES 报文(源描述 source description)所携带。

(3) 由于上述两项功能要求所有的会话参加者发送 RTCP 报文,因此为了能够在一个 RTP 会话中接纳更多的用户,RTCP 给出另外关于调整控制报文发送速率的方法。一个 RTP 用户能够通过接收来自于其他用户的 RTCP 报文,独立地了解会话参加者的数目,RTCP 利用该数值参数为每个会话参加者计算控制报文的发送间隔,以达到 RTCP 所占用的通信带宽不超过会话带宽 5%的目的。

(4) RTCP 提供了基本的会话控制功能,如在不同的会话参加者之间互换身份(identification)信息等。该项功能为可选项,主要用于控制较为松散的会话,这样的会话对用户的加入和离开没有严格的管理。由此可以看出,尽管 RTCP 为访问每一个会话参加者提供了便捷的手段,但它并未满足应用程序所需的全部通信控制要求,所以 RTP 允许应用在 RTCP 的基础上选用其它高层控制协议以满足其特定的应用需求。

其中(1)~(3)项功能是 RTCP 必备功能,而第四项为可选功能。

RTCP 报文的类型有下述多种:

SR (sender report): 发送者报告。

RR (receiver report): 接收者报告。

SDES (source description items): 源描述条款,包括 CNAME。

BYE: 指示与会者退出。

APP: 应用的特定功能。

每一个 RTCP 包都以和 RTP 数据报的固定部分类似的结构开始,后随的结构单元和 RTCP 包的类型相似,其长度随类型不同而不同,但都是 32 比特的整数倍。

3.3.2 SR 和 RR

RTP 报文的接收者可以利用两种类型的 RTCP 报告报文(SR 或 RR)来提供有关数据接收质量的统计信息,具体选用 SR 报文还是 RR 报文要看该接收者是否同时是一个 RTP 报文的发送者,明确地讲,如果一个会话参加者自最后一次发送 RTCP 报文后,发送了新的 RTP 数据报文,那么该参加者需要传送 SR 报文,否则传送 RR 报文。SR 报文和 RR 报文的主要区别在于前者包含了 20 字节有关发送者的信息。

SR 报文格式如图 3-3 所示。

SR 报文包括三部分,第一部分是 RTCP 报文头,长度为 8 个字节;第二部分是发送者信息,长度为 20 字节;第三部分为接收报告块,接收报告块携带了有关连续码流接收情况的统计信息,其中 SSRC 值表明了统计信息所针对的具体码流,在一个 RTP 会话中,参加者有可能收到多个具有相同媒体类型的码流,因而在一个 SR 或 RR 报文中可能会包含多个接收报告块。

SR 报文中的字段含义如下。

版本(V),填充(P)与 RTP 报头中的 V, P 字段含义相同。

RC: 5 个比特,表示包中包含的接收报告块的数目。

载荷类型(PT): 8 个比特,其值反映了该报文的类型(如对于 SR, PT 为 200;对于 RR, PT 为 201 等)。

长度(length): 16 个比特,表示 RTCP 包的长度,包括头和填充字节。

SSRC：32 个比特，该标识值一方面表明了该控制报文的产生者，另一方面反映了该报文携带的发送者信息块内的统计数据所针对的具体码流。

NTP 时间戳：64 个比特，它记录了 SR 报文的发送时刻，该值可被用来估测 RTP 报文或连续数据的平均发送速率。该时间戳为绝对时间戳，64 位的前 32 位是从 1900 年 1 月 1 日 0 时开始到现在的以秒为单位的整数部分，后 32 位是此时间的小数部分。



图 3-3 SR 报文格式

RTP 时间戳：32 个比特，同 NTP 时间戳相同，不过它的计时单位和计时原点与 RTP 数据报文的时间戳相同，该值可为码流内部或码流间的同步控制机制所选用。

发送报文计数：32 个比特，它记录了到该 SR 报文发送时所发的 RTP 数据报文的总数。

发送字节计数：32 个比特，它记录了到该 SR 报文发送时所发的连续数据的字节总数。

SSRC-n：同步源标识符，SR 中的信息是关于这些同步源的。

丢失百分比(fraction lost)：8 个比特，它表示自上一个 SR 或 RR 报文发送后，在数据接收过程中丢失的 RTP 报文数同所应收到的 RTP 报文总数的百分比。

包丢失累计数(cumulative number of packet lost)：24 个比特，它记录了到该 RTCP 报文发送时，在接收过程中所丢失的 RTP 数据报文的总数。

被扩展的最大次序号(extended highest sequence number)：32 个比特，由于

RTP 报文中的次序号只有 16 位，因而数据传输过程经过一段时间后，次序号将被回置为 0，RTP 协议称这段时间为一个次序号周期（sequence number cycle），相应于此，被扩展的最大次序号由两部分构成，低 16 位记录了到发送该 RTCP 报文时，收到的 RTP 报文的最大次序号值，高 16 位包含了所经历的次序号周期数，两者联合反映了应收到的 RTP 数据报文的总数。

包间抖动（IJ：interarrival jitter）：32 个比特，其所含内容是对数据传输过程中 RTP 报文所经历的延时抖动的一种估测，由下述两式确定。其中 S 为 RTP 报文的时间戳值，R 为 RTP 报文的到达时刻。IJ 值可以大致反映 RTCP 报文发送时网络的拥塞状况。

$$D(i-1,i)=(R_i-S_i)-(R_{i-1}-S_{i-1})$$

$$J=IJ+(|D(i-1,i)|-IJ)/16$$

上一个 SR（LSR：last SR timestamp）：32 个比特，其值为 NTP 时间戳的中间 32 位，表示最近收到 SR 报文的时间。

自从上一个 SR 的延时（DLSR）：32 个比特，其值为最近收到的 SR 报文到发送该 RTCP 报文的延时。

由此，该 RTCP 报文的接收者则可以通过 LSR 域和 DLSR 域来估算两点之间的往返传播延时，具体地说，若令 A 表示该 RTCP 报文的到达时刻，则连续数据流收/发两端之间的往返传播延时为 A-LSR-DLSR。

RR 报文格式如图 3-4 所示，它与 SR 大致相同，只不过不包含发送者信息块，因而不赘述。

0	8	16	31	
V	P	RC	PT=RR=201	长度(length)
发送者 SSRC				
SSRC-1				
丢失百分比		包丢失累计数		
最大扩展次序号				
包间抖动 (IJ)				
上一个 SR (LSR)				
自从上一个 SR 的时延 (DLSR)				
SSRC-2				
.....				
特定轮廓的扩展				

图 3-4 RR 报文格式

§ 3.4 轮廓文件和格式文件

对于一个具体应用而言，有关 RTP 协议的完整说明需要额外的一个或多个相关文件，这些文件可以分为两类：轮廓文件和格式文件。

由于 RTP 可以为许多类型的应用提供实时数据的传输服务，而这些应用又有着不同的需求及特点，因此为了增强其适应不同应用需求的灵活性，RTP 允许在一个独立的轮廓文件中，为在特定的应用环境中向特定类型的应用提供数据传输服务，就 RFC1889 给出的基本规定作恰当的调整和扩充。一般而言，RTP 要求不同类型的多媒体应用与不同轮廓文件之间有着一一对应的关系。

另一种类型的文件为格式文件，它定义了特定类型的负载数据（如符合 H.261 压缩标准的视频数据）在 RTP 报文中的具体格式，即定义了用户数据的打包方案。一个格式文件可为多个轮廓文件所引用，也可独立于任何轮廓文件而被具体应用单独使用。轮廓文件所包含的一个主要内容就是：定义 RTP 报文的负载类型值同格式文件之间的映射关系。

从实质上讲，轮廓文件连同其引用的格式文件是 RTP 服务同具体应用之间的一种约定。一方面，多媒体应用能够通过这种约定设置 RTP 的运行方式，如约定 RTP 报文某些域的取值含义、范围等，以此来满足其特定的应用需求；另一方面，RTP 服务能够借助于这些约定了解具体应用的特点，并实施某些特定的操作。

轮廓文件对 RTP 服务所能完成的调整和扩充可归纳为如下几个方面：

- RTP 报文的固定头：轮廓文件可以指定 RTP 报文头中标记域和负载类型域所占字节的比特分配方案，如允许标记域占用多个比特（缺省为 1）。

- 负载类型：轮廓文件中定义了 RTP 报文的负载类型值同格式文件的映射关系，以及 RTP 报文时间戳所采用的时钟频率，时钟频率的取值应当能够满足数据采样的精度要求。例如对于 H.261 视频流，RTP 时间戳频率为 90KHz。

- RTP 报头的扩展：轮廓文件可以选用两种扩充 RTP 报头的方式，一种方式是利用报头中的扩展位；另一种方式是在 RTP 固定头的 SSRC 域后直接添加有由轮廓文件定义的新域。相比较而言，后者开销少，且易于处理。

- RTCP 机制：轮廓文件或其引用的格式文件可以定义新的 RTCP 报文类型，指定 SR/RR 报文的扩充方式，以及 RTCP 调整报文所占会话带宽的比值、报文发送的最小间隔等参数。

- 轮廓文件要指定为 RTP 及 RTCP 分配传输层地址的方案，如规定 RTP 占用的 UDP 端口号为偶数而 RTCP 占用的 UDP 端口号为 RTP 的端口号加 1。

一般说来，轮廓文件对于 RTP 的调整及扩充反映了那些独立于数据负载编码类型的应用需求，而格式文件则根据连续数据的编码方式的特点对 RTP 的使用方式作出进一步的约定。