MP4格式分析

一、box介绍

mp4文件由一系列的box（也有人把它叫做atom）构成，每个box包含box头部和box体。box体可以包含普通的数据，也可以包含其他的box，如果box中包含了另一个box，这种box称为container box。

    box由head和body组成，head中指明了box的大小和类型，如果size为1，则表示这个box的大小为large size，真正的size值要在largesize域上得到。（实际上只有“mdat”类型的box才有可能用到large size）如果size为0，表示该box为文件的最后一个box，文件结尾即为该box结尾。（同样只存在于“mdat”类型的box中）

二、mp4中常见的box

1、ftyp类型的box。它是mp4文件的标识，包含了一些文件的信息

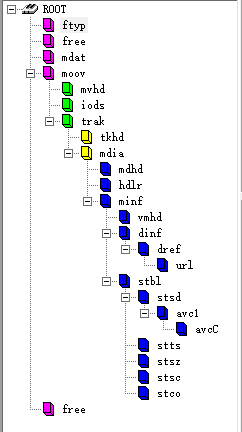
2、moov类型的box。它是一种container box，子box包含了媒体metadata信息

3、mdat类型的box。存放了媒体数据，它也是container box

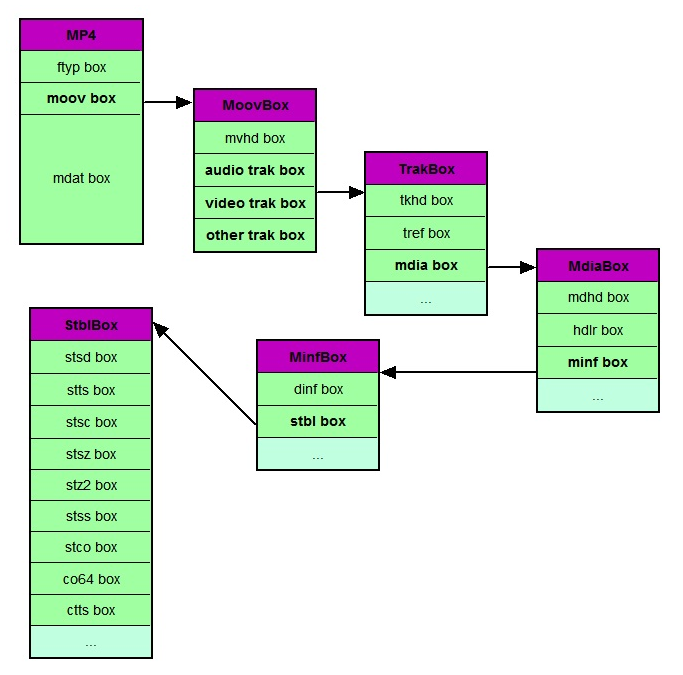
4、udta类型的box。存放用户自定义的信息

三、mp4中的一些概念  
    1、track。sample的集合  
    2、hint track。不包含数据，包含的是其他track组成的流媒体的指示信息。  
    3、sample。对于track来说是数据；对于hint track来说是流媒体包的格式  
    4、sample table。指明sample时域和物理布局的表  
    5、chunk。一个track中若干sample组成的集合

四、MP4示例图（用MP4info.exe可以查看生成的MP4文件结构）



五、常见的MP4结构图



六、具体box分析

    下面分析常见的box，他们的关系请参考上面的常见的MP4结构图。

ftyp

    ftyp box（File Type Box）。  
   有且只有一个，在mp4文件最开始的地方。  
    ftyp body依次包括1个32位的major brand（4个字符），1个32位的minor version（整数）和1个以32位（4个字符）为单位元素的数组compatible brands。这些都是用来指示文件应用级别的信息

moov

moov box（Movie Box）  
    该box包含了文件媒体的metadata信息，“moov”是一个container box，具体内容信息由子box诠释。同File Type Box一样，该box有且只有一个，且只被包含在文件层。一般情况下，“moov”会紧随“ftyp”出现。

一般情况下，“moov”中会包含1个“mvhd”和若干个“trak”。其中“mvhd”为header box，一般作为“moov”的第一个子box出现（对于其他container box来说，header box都应作为首个子box出现）。“trak”包含了一个track的相关信息，是一个container box

mvhd

mvhd box（Movie Header Box）

它是moov box的第一个子box

trak

trak box（Track Box）

它是moov box的子box，也是一个container box，其子box包含了该track的媒体数据引用和描述（hint track除外）。一个MP4文件中的媒体可以包含多个track，且至少有一个track，这些track之间彼此独立，有自己的时间和空间信息。“trak”必须包含一个“tkhd”和一个“mdia”，此外还有很多可选的box。其中“tkhd”为track header box，“mdia”为media box，该box是一个包含一些track媒体数据信息box的container box

tkhd

tkhd box（Track Header Box）

它是trak box的第一个子box

mdia

mdia box（Media Box）  
    “mdia”也是个container box，其子box的结构和种类比较复杂。

总体来说，“mdia”定义了track媒体类型以及sample数据，描述sample信息。一般“mdia”包含一个“mdhd”，一个“hdlr”和一个“minf”，其中“mdhd”为media header box，“hdlr”为handler reference box，“minf”为media information box。

mdhd

mdhd box（Media Header Box）

它是mdia box的第一个子box

hdlr

hdlr box（Handler Reference Box）

它是mdia box的子box。“hdlr”解释了媒体的播放过程信息，该box也可以被包含在meta box（meta）中。

minf

minf box（Media Information Box）

它是mdia box的子box。“minf”存储了解释track媒体数据的handler-specific信息，media handler用这些信息将媒体时间映射到媒体数据并进行处理。“minf”中的信息格式和内容与媒体类型以及解释媒体数据的media handler密切相关，其他media handler不知道如何解释这些信息。“minf”是一个container box，其实际内容由子box说明。

一般情况下，“minf”包含一个header box，一个“dinf”和一个“stbl”，其中，header box根据track type（即media handler type）分为“vmhd”、“smhd”、“hmhd”和“nmhd”，“dinf”为data information box，“stbl”为sample table box。

vmhd

vmhd box（Video Media Header Box）

它是minf box的第一个子box（即header box）。

hmhd

hmhd box（Hint Media Header Box）

它是minf box的第一个子box（即header box）。

nmhd

nmhd box（Null Media Header Box）

它是minf box的第一个子box（即header box）。非视音频媒体使用该box。

dinf

dinf box（Data Information Box）  
    它是minf box的子box。“dinf”解释如何定位媒体信息，是一个container box。“dinf”一般包含一个“dref”，即data reference box；“dref”下会包含若干个“url”或“urn”，这些box组成一个表，用来定位track数据。简单的说，track可以被分成若干段，每 一段都可以根据“url”或“urn”指向的地址来获取数据，sample描述中会用这些片段的序号将这些片段组成一个完整的track。一般情况下，当数据被完全包含在文件中时，“url”或“urn”中的定位字符串是空的。

“url”或“urn”都是box，“url”的内容为字符串（location string），“urn”的内容为一对字符串（name string and location string）。当“url”或“urn”的box flag为1时，字符串均为空。

stbl

stbl box（Sample Table Box）

它是minf box的子box。“stbl”几乎是普通的MP4文件中最复杂的一个box了，首先需要回忆一下sample的概念。sample是媒体数据存储的单位，存储在media的chunk中，chunk和sample的长度均可互不相同。

“stbl”包含了关于track中sample所有时间和位置的信息，以及sample的编解码等信息。利用这个表，可以解释sample的时序、类型、大小以及在各自存储容器中的位置。“stbl”是一个container box，其子box包括：sample description box（stsd）、time to sample box（stts）、sample size box（stsz或stz2）、sample to chunk box（stsc）、chunk offset box（stco或co64）、composition time to sample box（ctts）、sync sample box（stss）等

stsd

stsd box（Sample Description Box）

它是stbl box的子box。“stsd”必不可少，且至少包含一个条目，该box包含了data reference box进行sample数据检索的信息。没有“stsd”就无法计算media sample的存储位置。“stsd”包含了编码的信息，其存储的信息随媒体类型不同而不同box header和version字段后会有一个entry count字段，根据entry的个数，每个entry会有type信息，如“vide”、“sund”等，根据type不同sample description会提供不同的信息，例如对于video track，会有“VisualSampleEntry”类型信息，对于audio track会有“AudioSampleEntry”类型信息。

视频的编码类型、宽高、长度，音频的声道、采样等信息都会出现在这个box中。

stts

stts box（Time To Sample Box）

它是stbl box的子box。“stts”存储了sample的duration，描述了sample时序的映射方法，我们通过它可 以找到任何时间的sample。“stts”可以包含一个压缩的表来映射时间和sample序号，用其他的表来提供每个sample的长度和指针。表中每 个条目提供了在同一个时间偏移量里面连续的sample序号，以及samples的偏移量。递增这些偏移量，就可以建立一个完整的time to sample表

stsz

stsz box（Sample Size Box）

它是stbl box的子box。“stsz” 定义了每个sample的大小，包含了媒体中全部sample的数目和一张给出每个sample大小的表。这个box相对来说体积是比较大的。

stsc  
    stsc box（Sample To Chunk Box）

    它是stbl box的子box。用chunk组织sample可以方便优化数据获取，一个thunk包含一个或多个sample。“stsc”中用一个表描述了sample与chunk的映射关系，查看这张表就可以找到包含指定sample的thunk，从而找到这个sample。

stss

stss box（Sync Sample Box）

它是stbl box的子box。stss”确定media中的关键帧。对于压缩媒体数据，关键帧是一系列压缩序列的开始帧，其解压缩 时不依赖以前的帧，而后续帧的解压缩将依赖于这个关键帧。“stss”可以非常紧凑的标记媒体内的随机存取点，它包含一个sample序号表，表内的每一 项严格按照sample的序号排列，说明了媒体中的哪一个sample是关键帧。如果此表不存在，说明每一个sample都是一个关键帧，是一个随机存取 点。

stco

    stco box（Chunk Offset Box）  
    它是stbl box的子box。stco”定义了每个thunk在媒体流中的位置。位置有两种可能，32位的和64位的，后者对非常 大的电影很有用。在一个表中只会有一种可能，这个位置是在整个文件中的，而不是在任何box中的，这样做就可以直接在文件中找到媒体数据，而不用解释 box。需要注意的是一旦前面的box有了任何改变，这张表都要重新建立，因为位置信息已经改变了。

free

    free或skip box（Free Space Box）  
    “free”中的内容是无关紧要的，可以被忽略。该box被删除后，不会对播放产生任何影响。

mdat

mdat box（Meida Data Box）

该box包含于文件层，可以有多个，也可以没有（当媒体数据全部为外部文件引用时），用来存储媒体数据。数据直接跟在box type字段后面，具体数据结构的意义需要参考metadata（主要在sample table中描述）。