

SMPTE ST 2110-10: IP Network Fundamentals (Repost)

This article is the first in a series describing some key technical features of each part of the new SMPTE ST 2110:2017 "Professional Media Over Managed IP Networks" standard.

At the time of writing, the SMPTE 2110 standard comprises multiple parts, with more to be added in the future. The first three parts are:

- ST 2110-10 System Timing and Definitions
- ST 2110-20 Uncompressed Active Video
- ST 2110-30 PCM Digital Audio

Several parts on video traffic shaping and ancillary data will be added in the near future.

To begin this series, ST 2110-10 is an excellent starting point. We begin by defining a series of common elements applicable to the remaining standards in the 2110 family.

Specifically, ST 2110-10 defines the transport layer protocol, datagram size limitations, Session Description Protocol (SDP) requirements, and clocks along with their timing relationships.

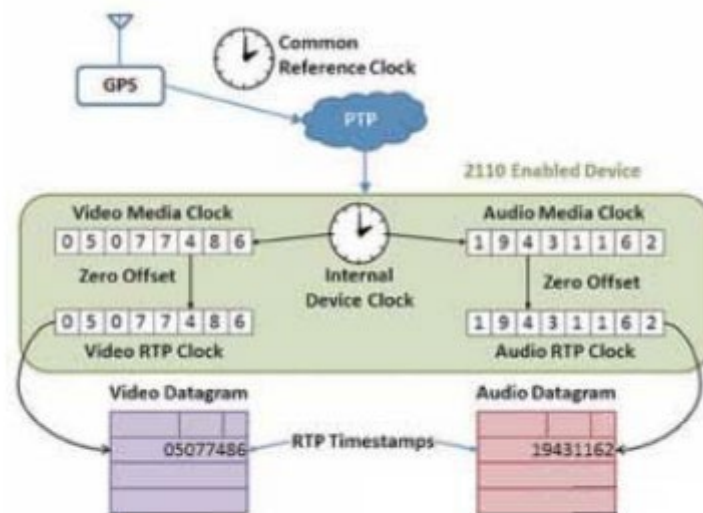


图1 ST 2110-10中定义的时钟

2110 Datagrams

The Real-time Transport Protocol (RTP, as defined in IETF RFC 3550) was chosen for 2110 applications because it contains many features required for transmitting media signals over IP networks, while lacking some undesirable aspects of TCP (Transmission Control Protocol). Using RTP, uncompressed media samples are segmented into datagrams, each prefixed with both an RTP header and an additional UDP header before being passed to the IP layer for further processing. ST 2110-10 defines a standard 1460-byte UDP payload size limitation (including UDP and RTP headers),

which provides sufficient space for over 450 24-bit audio samples or approximately 550 pixels of 4:2:2 10-bit uncompressed video signals. 2110 also defines an extended UDP size limitation of 8960 bytes, which may be useful on networks supporting Ethernet jumbo frames.

Timing is Everything

Perhaps the most critical aspect of ST 2110-10 relates to clocks and timing. Unlike SDI and MPEG transport streams, each media type in a 2110 system is transmitted as an independent IP packet stream. This necessitates a mechanism to realign them according to their correct temporal relationships once the signals traverse a network. Figure 1 shows several different clocks defined in this specification; each clock serves a specific purpose:

The Precision Time Protocol (PTP) is used to distribute a common reference clock (accurate to within 1 ms) to every device on the network; Each signal source maintains an internal device clock, which can (and should) be synchronized to the common reference clock; Each signal type corresponds to a media clock advancing at a fixed rate, associated with the frame rate or sampling rate of each media signal; An RTP clock is used within each signal source to generate the RTP timestamps included in the header of each RTP datagram for that media signal. The RTP clocks are synchronized to their respective media clocks.

To understand how these clocks relate, consider an audio source sampled at 48 kHz. Within that audio source (e.g., a microphone network adapter), an internal device clock will run, providing the overall timebase for this device. This clock should be synchronized to the accurate common reference clock delivered via PTP over the network, which in most cases will itself be synchronized to a GPS receiver. The internal device clock is then used to establish a media clock running accurately at 48 kHz, so each "tick" of the media clock counter will correspond to an audio sampling instant. The RTP clock used for the audio stream will then be directly related to the media clock and used to generate the RTP timestamps within each datagram's header.

The RTP timestamps used in 2110 systems represent the actual sampling time of the media signal contained within each datagram. For an audio datagram (typically containing more than one sample per channel), the RTP timestamp represents the sampling time of the first audio sample contained in the datagram. For video datagrams (typically requiring multiple datagrams to transmit each video frame), an RTP timestamp may be shared by hundreds or even thousands of datagrams. Note that individual datagrams will have unique packet sequence numbers, allowing them to be distinguished from each other and used for purposes such as packet loss detection and error correction.

As datagrams traverse the IP network, they will typically encounter varying amounts of transmission delay. This may arise from buffering within IP network devices or potentially from

different processing measures applied to various media types. At their destination, the timestamps within the RTP datagram headers can be used to align all the different media samples onto a common timeline, allowing individual audio and video signals to be correctly synchronized for processing, playout, or storage.

The ability to precisely synchronize multiple media types at any point in a broadcast chain provides one of the primary values of 2110.

It allows each media stream to be processed independently. This means a device performing audio level normalization no longer needs to de-embed audio from an SDI video stream, process the audio, and then re-embed it back into SDI. Instead, each media signal can follow its own path through the processing steps required only for that type of media, and still be resynchronized to other media when necessary. Furthermore, if all media clocks are referenced to a GPS-sourced clock, then via 2110, two cameras hundreds of meters or even hundreds of kilometers apart can be genlocked together without needing to send a sync signal from one to the other. This is a very cool thing. Subsequent articles will discuss SDP (Session Description Protocol).