Cimarron Systems

RTP Protocol Transport of H.264 Video and MPEG I/II Layer 3 Audio

Application Note: AN104

May 4, 2018

Cimarron Systems, LLC

Table of Contents

List of Figures

Document: AN-180504-1

Document Status: Released

Revision: 1.0

Using the RTP Protocol to Transport Video and Audio

This application note describes the use of both the Real-time Transport Protocol (RTP) and the Real-time Transport Control Protocol to simultaneously transport H.264 video and MPEG-I/II Layer 3 (MP3) audio bitstreams across the network. Specifically, the RTP Server / Client Applications—available as a component of several Cimarron Systems, LLC Digital Media Software Development Kit (DMSDK) products—work jointly to implement these protocols.

DMSDK Development Environment

Figure 1 shows a context diagram of the typical development environment in which the DMSDK operates, including: the NXP / Freescale iMX6 (or TI TMS320DM36x) platform running a number of the DMSDK components, a Ubuntu Linux Host computer, an audio / video source, and a video display / audio output device.

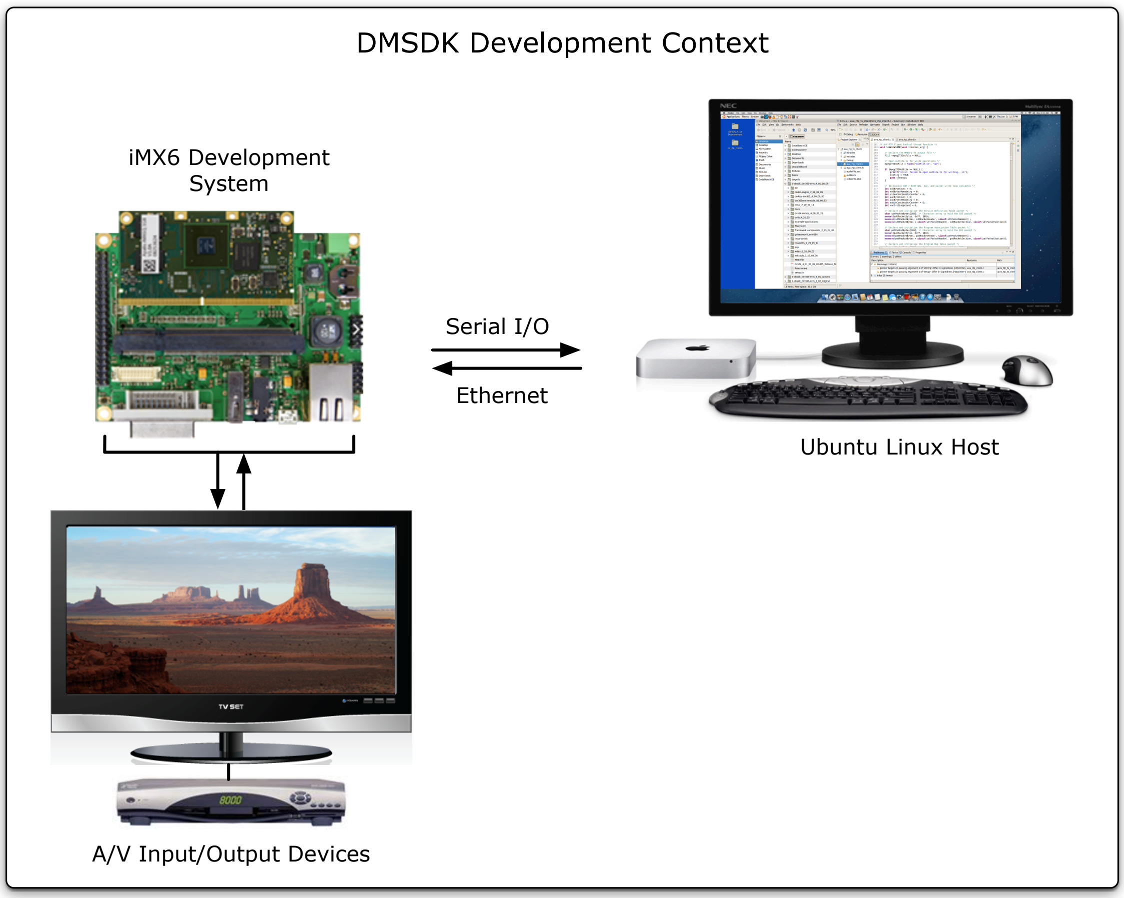

Figure 1: Digital Media SDK Development Context Diagram.

RTP Client/Server Application Architecture

As illustrated in Figure 2, RTP Server and RTP Client Application include: an RTP Video Transmitter for transport of H.264 encoded video; an RTP Audio Transmitter for transport of MPEG-I/II Layer 3 encoded audio; an RTCP Server for transmission as well as reception of RTCP Reception Reports and RCTP Receiver Bye messages; and an RTCP Client for transmission of Reception Reports as well as reception of RTCP Sender Reports and RTCP Sender Bye messages.

The RTP Server / Client Applications implement jointly H.264 video streaming and MPEG I/II Layers 3 audio streaming in accordance with IETF *RFC 3550 RTP: A Transport Protocol for Real-Time Applications*, *RFC 6184 RTP Payload for H.264 Video*, *RFC 3551 RTP Profile for Audio and Video Conferences with Minimal Control*, and *RFC 2250 RTP Payload Format for MPEG1 / MPEG2 Video*.

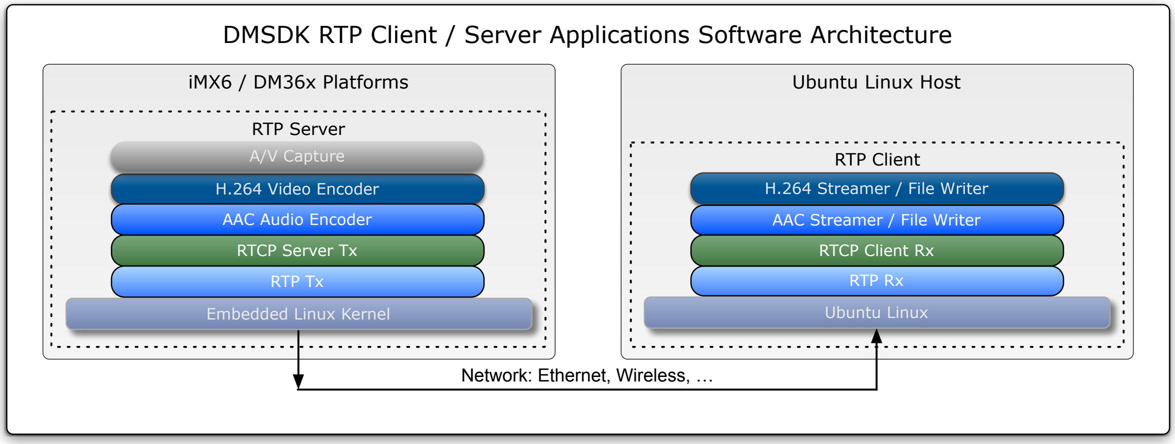

Figure 2: RTP Client / Server Application Architecture.

RTP Protocol Transport of H.264 Video / MP3 Audio

Figure 3 shows the structure of the RTP H.264 Video Common Packet—constructed in accordance with IETF *RFC 6184 RTP Payload for H.264 Video*—which implements a simple method to signal the start of each H. 264 Network Abstraction Layer (NAL) Unit as well as its Raw Byte Sequence Payload (RBSP), i.e., the NAL Unit bitstream data bytes (as detailed below, the NAL Unit start signaling method appends a single byte to the end of the standard RTP Packet Header).

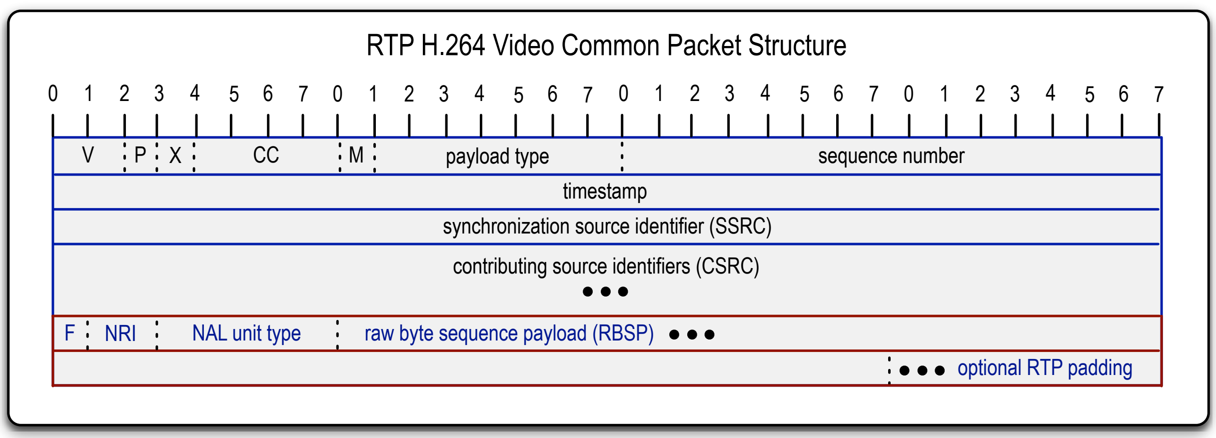

Figure 3: RTP H.264 Video Common Packet Structure.

Figure 4 shows the structure of the RTP MPEG-I/II Layer 3 Audio Packet—constructed in accordance with IETF *RFC 3550 RTP: A Transport Protocol for Real-Time Applications*—which implements carriage of MPEG-I/II Layer 3 audio bitstreams.[1] [2]

---

[1] See the *iMX6-NF, CS365-TI, CS36x-TI, and/or CS36x-LI DMSDK Datasheets* for details regarding the Cimarron Systems, LLC development system environment, features, architecture, and capabilities.

[2] For an excellent tutorial concerning the application of RTP, please see the presentation *RTP: Multimedia Streaming over IP*, by Colin Perkins, USC Information Sciences Institute.

As described in RFC 6184, large H.264 NAL Units may need to be fragmented into packets less than or equal to the size of one MTU—typically 1500 bytes but possibly smaller, depending on the transmission network—as a result, large NAL Units are fragmented for transmission by the RTP Server using RTP Type A fragmentation units (FU-A) then reassembled by the RTP Client at the receiver.

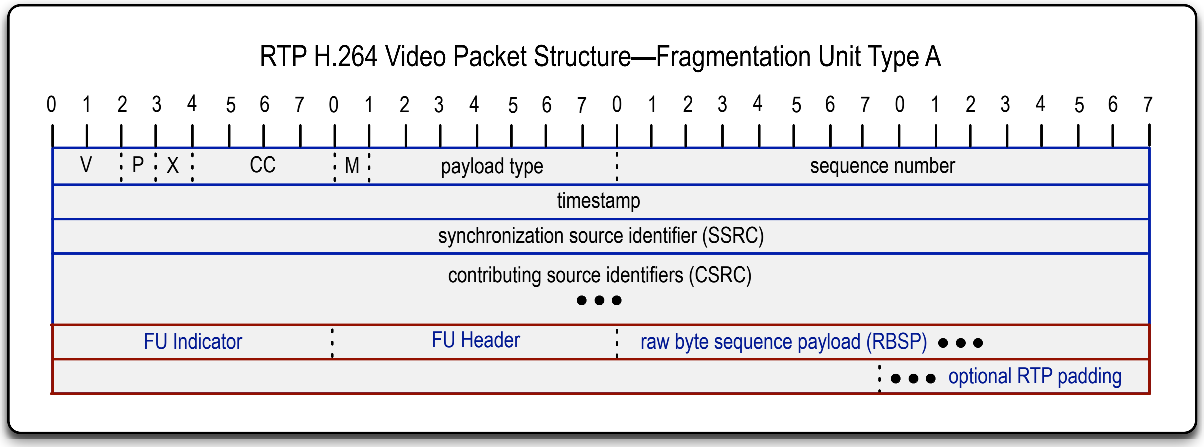

Figure 4: RTP H.264 Fragmentation Unit Type A Packet Structure.

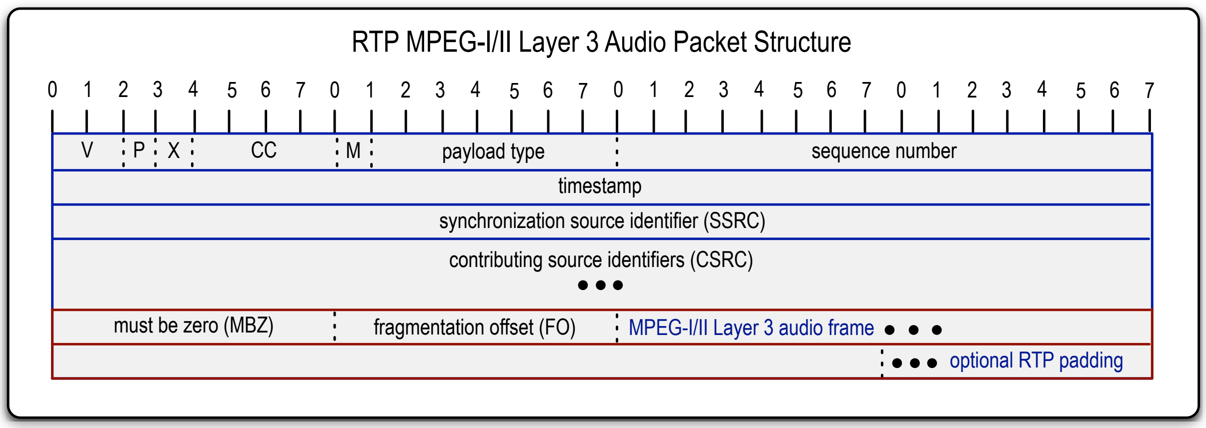

Figure 5: RTP MPEG-I/II Layer 3 Audio Packet Structure.

Common RTP Packet Syntax

- **V:** (2 bits) RTP Protocol Version, always equal to 2.
- **P:** (1 bit) If the Padding bit is set, the packet contains one or more additional padding octets at the end which are not part of the payload.
- **X:** (1 bit) If the Extension bit is set, the fixed header MUST be followed by exactly one Header Extension.
- **CC:** (4 bits) The CSRC Count contains the number of CSRC identifiers that follow the fixed header.
- **M:** (1 bit) The interpretation of the Marker is defined by a profile—it is intended to allow significant events such as frame boundaries to be marked in the packet stream.
- **Payload Type:** (7 bits) This field identifies the format of the RTP payload and determines its interpretation by the application.
- **Sequence Number:** (16 bits) The Sequence Number increments by one for each RTP data packet sent and may be used by the receiver to detect packet loss and to restore packet sequence

(the initial value of the sequence number should be random).

- **Timestamp:** (32 bits) The Timestamp reflects the sampling instant of the first octet in the RTP data packet. The sampling instant MUST be derived from a clock that increments monotonically and linearly in time to allow synchronization and jitter calculations.
- **SSRC:** (32 bits) The SSRC field identifies the Synchronization Source—this identifier should be chosen randomly, with the intent that no two Synchronization Sources within the same RTP session will have the same SSRC Identifier.
- **CSRC:** (32 bits) The CSRC List identifies the contributing sources for the payload contained in this packet. The number of identifiers is given by the CC Field—only 15 can be identified—not used in the present implementation.
- **Padding:** (M bytes) Padding bytes added to the end of a packet in order to force an whole number of packet bytes.

H.264 RTP Single NAL Unit Packet Syntax

- **F:** (1 bit) A value of 0 indicates that the NAL Unit Type octet and payload should not contain bit errors or other syntax violations. A value of 1 indicates that the NAL Unit Type octet and payload may contain bit errors or other syntax violations.
- **NRI:** (2 bits) A value of 00 indicates that the content of the NAL Unit is not used to reconstruct reference pictures for inter picture prediction—value greater than 00 indicate that the decoding of the NAL Unit is required to maintain the integrity of the reference pictures.
- **NAL Unit Type:** (5 bits) Interpretation of NAL Unit Type values 0-23 are the same as ITU-T H.264 recommendation—see RFC 6184 Table 3 for the interpretation of values 24-29.
- **Raw Byte Sequence Payload:** (N bytes) A sequence of bytes constituting the H.264 video NAL bitstream.

H.264 RTP Fragmentation Unit Packet Syntax

- **FU Indicator:** (8 bits) As described in Section 5.3 of RFC 6184, contains the normal NAL Unit **F**, **NRI** bit as well as the **Type** (5 bits) as described in Table 2.
- **FU Header:** (8 bits) As described in Section 5.8 of RFC 6184, contains the **S** set to '1' indicating the start of the fragmentation unit '0' otherwise, the **E** bit set to '1' indicating the end of the fragmentation unit ''0' other, and the **Type** (5 bits) as described in Table 7.1 of ITU Recommendation H.264, *Advanced video coding for generic audiovisual services*, March 2010.

MPEG-I/II Layer 3 RTP Packet Syntax

- **Must Be Zero:** (8 bits) Must Be Zero byte (MBZ), as specified in RFC 2250.
- **Fragmentation Offset:** (8 bits) Fragmentation Offset, (FO) byte, as specified in RFC 2250.
- **Audio Data Transport Stream** (1154 bytes per frame (576 bytes for joint stereo): A sequence of bytes constituting the MPEG-I/II Layer 3 audio bitstream.

RTCP QoS Messages

- **Sender Report:** Sender SSRC, NTP Timestamp, Packet Count, and Octet Count.
- **Receiver Report:** ReceiverSSRC, Cumulative Number of Packets Lost, Fraction of Packets Lost, Extended Highest Sequence Number, Inter-arrival Jitter, Last Sequence Number, Delay Since Last Sender Report, and Source Description Items.
- **Sender Bye:** Sender Bye message sent when the RTP Server ends transmission.
- **Receiver Bye:** Receiver Bye message sent when RTP Server ends its transmission, the receiver leaves the session, or the connection is poor or lost.

RTP H.264 Video/MPEG I/II Layer 3 Audio Transport Timing

The timing diagram shown in Figure 5 illustrates the relationship between audio/video sample frame timing, packet transmission timing, and RTCP packet transmission / reception for the RTP Client / Server Applications when set for H.264 (30 fps) / MPEG-I/II Layer 3 (32 ks/s). In the present example, the video frame rate is 30 Hz while the audio sampling rate is 32 kHz; however, both RTP timestamps are based on the same 90 kHz clock.

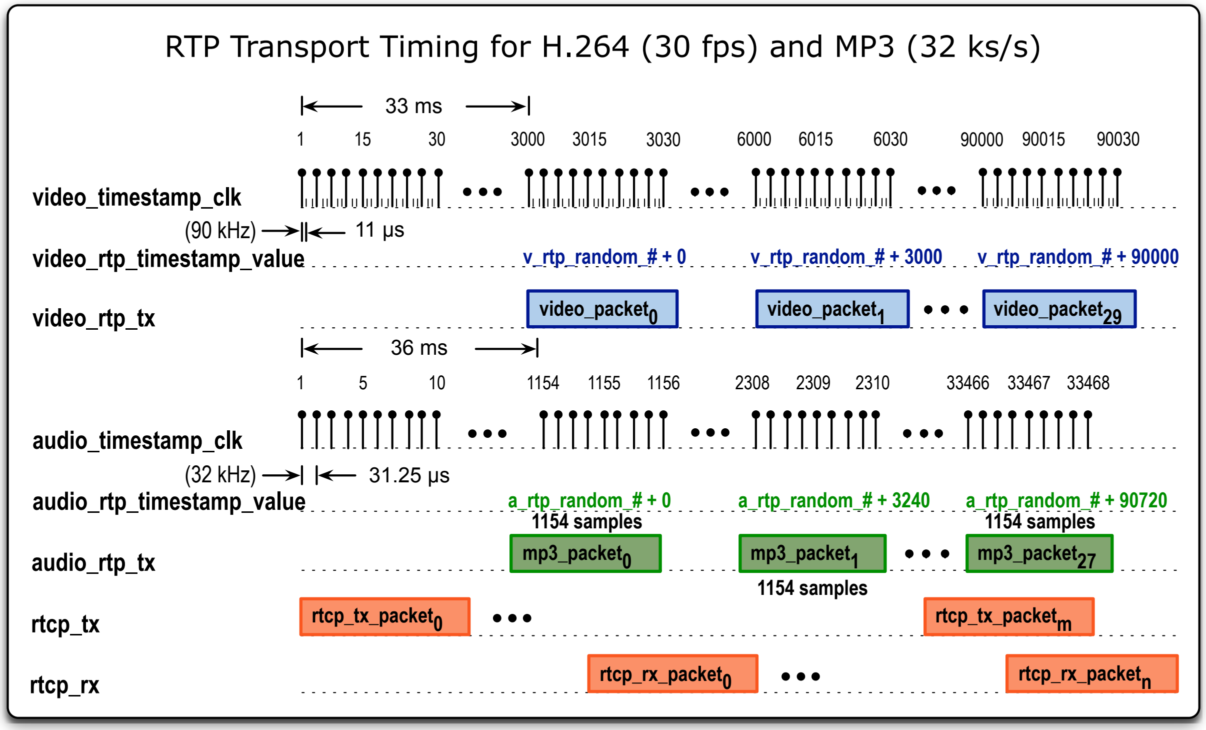

Figure 6: RTP Protocol H.264 Video / MPEG-I/II Layer 3 Audio Transport Timing Diagram.

RTP Client/Server Application Functional Diagram

Figure 7 shows a functional block diagram (blocks in green) of the RTP Client / Server Application portion of the software architecture illustrated above in Figure 2. (Note: that the blue colored blocks represent the functional components that starts up / shuts down the application; reads the keyboard/mouse inputs; initializes the audio / video codecs; captures and encodes the audio / video frames; constructs and transmits the audio / video RTP packets; composes then transmits RTCP Sender Report messages; receives then interprets the RTCP Receiver Report messages; and, based upon QoS status, adjusts application performance parameters as required).

RTP Server

- **main():** The main() thread creates the three additional threads; i.e., the rtcpServer(), videoTransmitter(), and audioTransmitter() then, after the threads each complete, terminates them in an orderly fashion.
- **rtcpServer():** The rtcpServer() thread creates / sends RTP Video and Audio Sender Reports as well as reads / responds to RTP Video and Audio Receiver Reports. Finally, the rtcpServer() thread receives Video and Audio Reception Bye messages then takes appropriate actions.
- **videoTransmitter():** The videoTransmitter() thread reads H.264 NAL Units from the encoder / file, subsequently forms the video RTP packets, then transmits them out to port 5004.

- **audioTransmitter():** The audioTransmitter() thread reads the MPEG I/II Layer 3 encoded audio

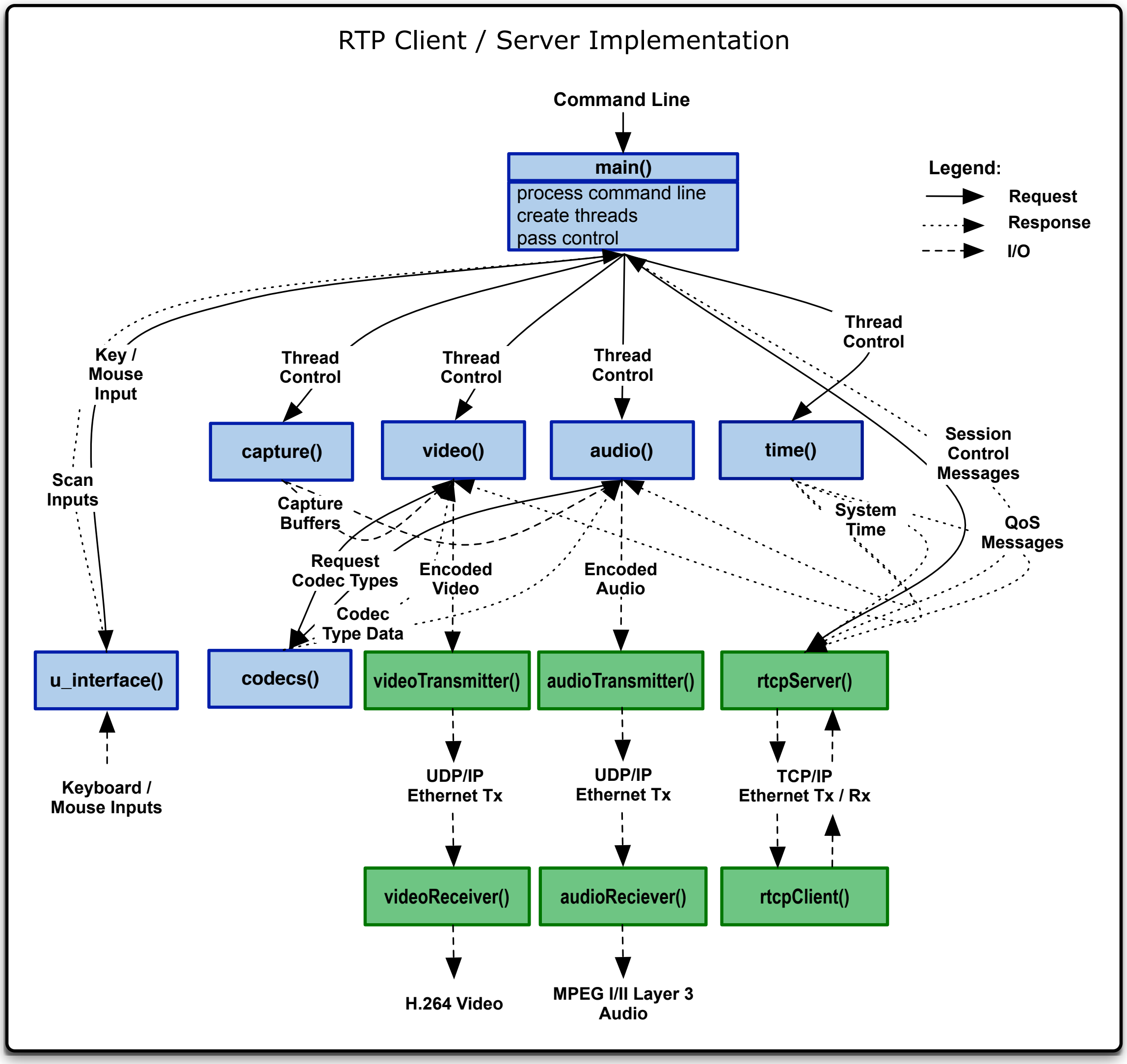

Figure 7: RTP Client / Server Application Software Functional Block Diagram.

frames from the encoder / file, subsequently forms the audio RTP packets, then transmits them out to port 5006.

RTP Client

- **rtcpClient():** The rtcpClient() thread computes QoS statistics from received video and audio packets, creates / sends RTP Video and Audio Reception Reports. The rtcpClient() thread also receives / accounts for RTP Video and Audio Sender Reports. Finally, the rtcpClient() thread receives Video and Audio Sender Bye messages then takes appropriation actions.
- **videoReceiver():** The videoTransmitter() thread reads incoming RTP H.264 NAL Unit packets from port 5004, then transmits the depacketized bitstream either to a player application or to a file.
- **audioReceiver():** The audioTransmitter() thread reads incoming RTP MPEG I/II Layer 3 packets from port 5006, then transmits the depacketized bitstream either to a player application or to a file.

RTP Client/Server Session Wireshark Capture

Figure 7 shows details of a typical RTP Client / Server H.264 Video / MPEG I/II Layer 3 Audio transmission session (Note: The source Wireshark capture files are available upon request by sending an email to info@cimarronsystems.com. Please put "RTP Client / Server Wireshark captures…" in the subject line.

No.	Time	Protocol	Info
44565	240.105839	RTP	PT=MPEG-I/II Audio, SSRC=0x68021FF, Seq=6565, Time=21270600, Mark
44566	240.127336	RTP	PT=DynamicRTP-Type-96, SSRC=0x5CECEE93, Seq=32020, Time=21309000
44567	240.127529	RTP	PT=DynamicRTP-Type-96, SSRC=0x5CECEE93, Seq=32021, Time=21309000
44568	240.127536	RTP	PT=DynamicRTP-Type-96, SSRC=0x5CECEE93, Seq=32022, Time=21309000, Mark
44569	240.141562	RTP	PT=MPEG-I/II Audio, SSRC=0x68021FF, Seq=6566, Time=21273840, Mark
44570	240.160549	RTP	PT=DynamicRTP-Type-96, SSRC=0x5CECEE93, Seq=32023, Time=21312000
44571	240.160747	RTP	PT=DynamicRTP-Type-96, SSRC=0x5CECEE93, Seq=32024, Time=21312000
44572	240.160754	RTP	PT=DynamicRTP-Type-96, SSRC=0x5CECEE93, Seq=32025, Time=21312000, Mark
44573	240.177743	RTP	PT=MPEG-I/II Audio, SSRC=0x68021FF, Seq=6567, Time=21277080, Mark
44574	240.184809	RTCP	Sender Report
44575	240.190374	RTCP	Receiver Report Source description

Figure 8a: RTP Client / Server H.264 Video / MPEG I/II Layer 3 Session Packet Sequence.

```
▼ Real-time Transport Control Protocol (Sender Report)
    10.. .... = Version: RFC 1889 Version (2)
    ..0. .... = Padding: False
    ...0 0000 = Reception report count: 0
    Packet type: Sender Report (200)
    Length: 6 (28 bytes)
    Sender SSRC: 0x5cecee93 (1559031443)
    Timestamp, MSW: 3734454727 (0xde9741c7)
    Timestamp, LSW: 730410976 (0x2b892fe0)
    [MSW and LSW as NTP timestamp: May  4, 2018 20:32:07.170062057 UTC]
    RTP timestamp: 21315000
    Sender's packet count: 32026
    Sender's octet count: 30077741
  [RTCP frame length check: OK - 28 bytes]
```

Figure 8b: RTCP H.264 Video Sender Report.

```
▼ Real-time Transport Control Protocol (Receiver Report)
    10.. .... = Version: RFC 1889 Version (2)
    ..0. .... = Padding: False
    ...0 0001 = Reception report count: 1
    Packet type: Receiver Report (201)
    Length: 7 (32 bytes)
    Sender SSRC: 0x9426307d (2485530749)
  ▼ Source 1
      Identifier: 0x5bebed93 (1542188435)
    ▼ SSRC contents
        Fraction lost: 0 / 256
        Cumulative number of packets lost: 0
    ▼ Extended highest sequence number received: 119673
        Sequence number cycles count: 1
        Highest sequence number received: 54137
      Interarrival jitter: 4513
      Last SR timestamp: 1114203971 (0x42696743)
      Delay since last SR timestamp: 338 (5 milliseconds)
```

Figure 8c: RTCP H.264 Video Receiver Report.


```
▼ Real-time Transport Control Protocol (Source description)
    10.. .... = Version: RFC 1889 Version (2)
    ..0. .... = Padding: False
    ...0 0001 = Source count: 1
    Packet type: Source description (202)
    Length: 7 (32 bytes)
  ▼ Chunk 1, SSRC/CSRC 0x9426307D
      Identifier: 0x9426307d (2485530749)
    ▼ SDES items
        Type: CNAME (user and domain) (1)
        Length: 19
        Text: cimarronsystems.com
        Type: END (0)
  [RTCP frame length check: OK - 64 bytes]
```

Figure 8d: RTCP MPEG I/II Layer 3 Audio Sender Source Description Report.

```
▼ Real-time Transport Control Protocol (Goodbye)
    10.. .... = Version: RFC 1889 Version (2)
    ..0. .... = Padding: False
    ...0 0001 = Source count: 1
    Packet type: Goodbye (203)
    Length: 1 (8 bytes)
    Identifier: 0x068021ff (109060607)
  [RTCP frame length check: OK - 8 bytes]
```

Figure 8e: RTCP MPEG I/II Layer 3 Audio Sender Bye.

Additional Resources

In addition to those already described, listed below are a number of resources that may be helpful:

1. RTP FAQ at Columbia University
2. RTP Tools at Columbia University
3. WireShark Capture for RTP

For more information regarding this and other Cimarron Systems, LLC products or to provide comments regarding this document, please contact us using the contact information below.

Contact Information:

Cimarron Systems, LLC
Evergreen, Colorado
telephone: +1 (303) 674-9207
email: info@cimarronsystems.com
https://www.cimarronsystems.com

Revision History:

Date	Version	Notes
5/4/2018	version 1.0	Initial version

