Cimarron Systems

RTP Protocol Transport of H.264 Video and AAC Audio

Application Note: AN100

November 15, 2012

Cimarron Systems, LLC

Table of Contents

List of Figures

Document**:** AN-121115-1
Document Status: Released
Revision: 1.0

Using the RTP Protocol to Transport Video and Audio

The present application note describes the use of the RTP Internet Protocol (IP) to simultaneously transport H.264 video and AAC audio bitstreams across the network. Specifically, the RTP A/V Streaming Client/ Server Applications—available as part of the Cimarron Systems Digital Media Software Development Kits (DMSDK)—implements many of the concepts outlined here.

DMSDK Development Environment

Figure 1 shows a context diagram of the typical development environment in which the DMSDK operates, including: the TM320DM36x DVEVM 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: DMSDK Development Context Diagram.

RTP A/V Streaming Client/Server Application Architecture

As illustrated in Figure 2, RTP A/V Streaming Server and RTP A/V Streaming Client Applications include: a RTCP Transmitter User Agent; an A/V Encoder set for encoding video to H.264 and audio to AAC-LC; and a RTP Transmitter. The Streaming A/V Decoder component includes: a RTP Receiver; a RTCP Receiver User Agent; and a MPEG-2 Transport Stream (TS) Encoder that multiplexes the encoded H.264 video and AAC-LC audio into a transport stream.

Both the RTP A/V Streaming Server and RTP A/V Streaming Client Applications implement H.264 video and AAC 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 3550 RTP: A Transport Protocol for Real-Time Applications, and RFC 3551 RTP Profile for Audio and Video Conferences with Minimal Control.

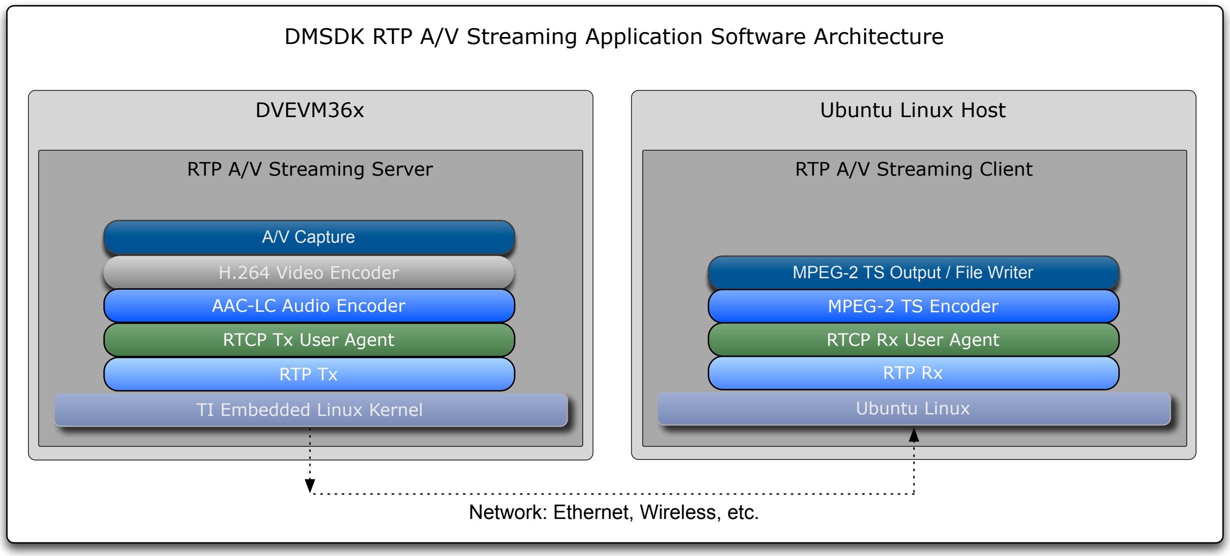

Figure 2: RTP A/V Streaming Client/Server Applications.

RTP Protocol Transport of H.264 Video / AAC Audio

Figure 3 shows the structure of the RTP H.264 Video 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).

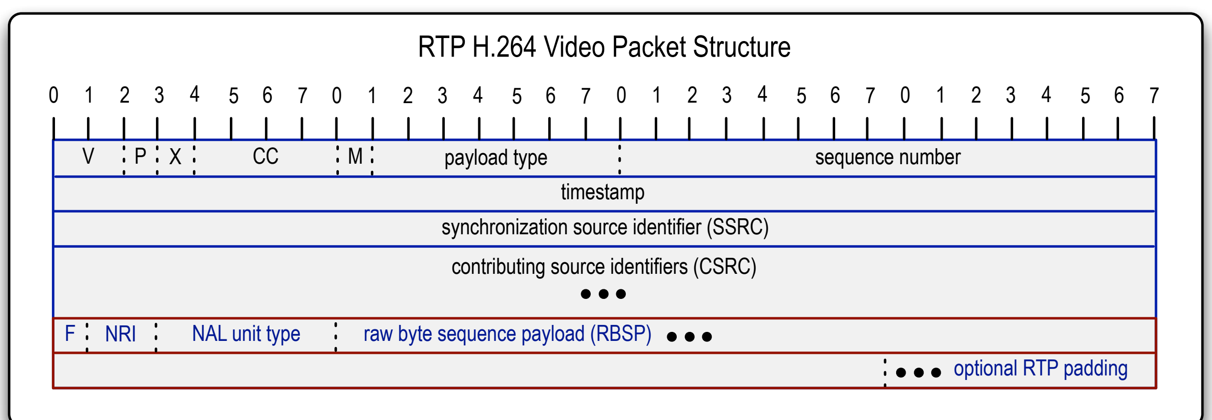

Figure 3: RTP H.264 Video Packet Structure.

Figure 4 shows the structure of the RTP AAC Audio Packet—constructed in accordance with IETF RFC 3550 *RTP: A Transport Protocol for Real-Time Applications*—which implements carriage of the AAC audio bitstream via the common Audio Data Transport Stream (ADTS).[1 2]

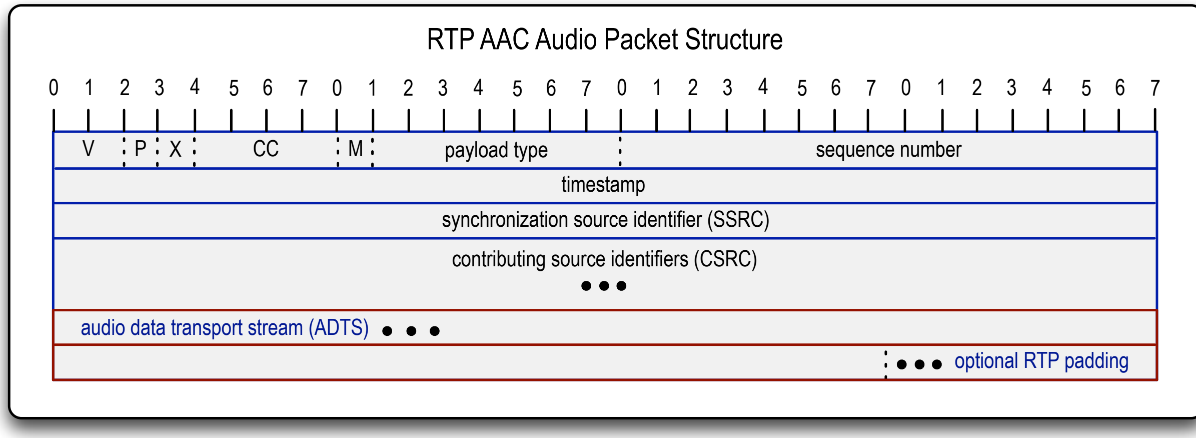

Figure 4: RTP AAC Audio Packet Structure.

The syntax descriptions present below apply to both RTP H.264 Video and AAC Audio Packets:

- **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.

---

[1] See the *CS365-TI DMSDK* and/or *CS368-TI Datasheets* for details regarding Cimarron Systems DMSDK development 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.

The following syntax descriptions apply only to RTP H.264 Video Packets:

- **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.

The following syntax descriptions apply only to RTP AAC Audio Packets:

- **Audio Data Transport Stream** (1024 bytes): A sequence of bytes constituting the AAC audio bitstream.

RTP H.264 Video/AAC 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 A/V Client/Streaming Server Applications when set for H.264 (30 fps)/AAC (32 ks/s). In the present example, the video timestamp clock is 90 kHz while the audio timestamp is 32 kHz, i.e., the audio sampling rate.

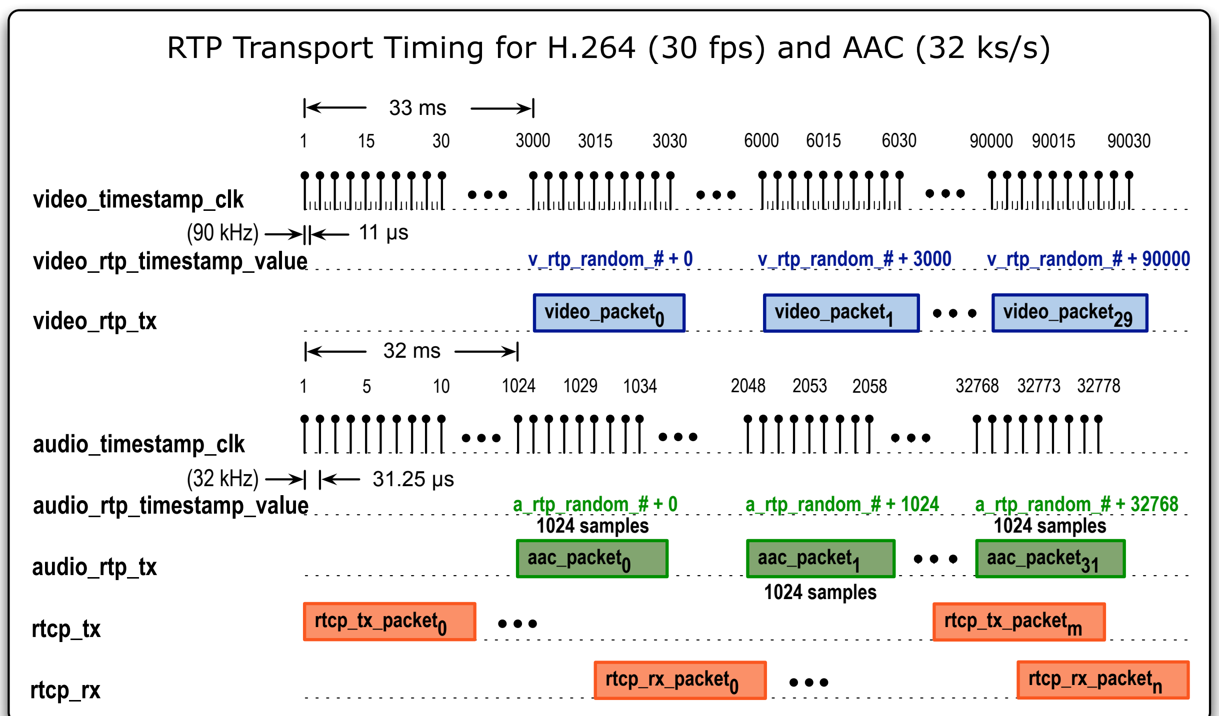

Figure 5: RTP Protocol H.264 Video/AAC Audio Transport Timing Diagram.

RTP A/V Streaming Server Application Functional Diagram

Figure 6 shows a functional block diagram of the RTP A/V Streaming Server Application portion of the software architecture illustrated 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 Session Control Messages; receives then interprets the RTCP QoS Messages; and, based upon QoS status, adjusts application performance parameters as required.

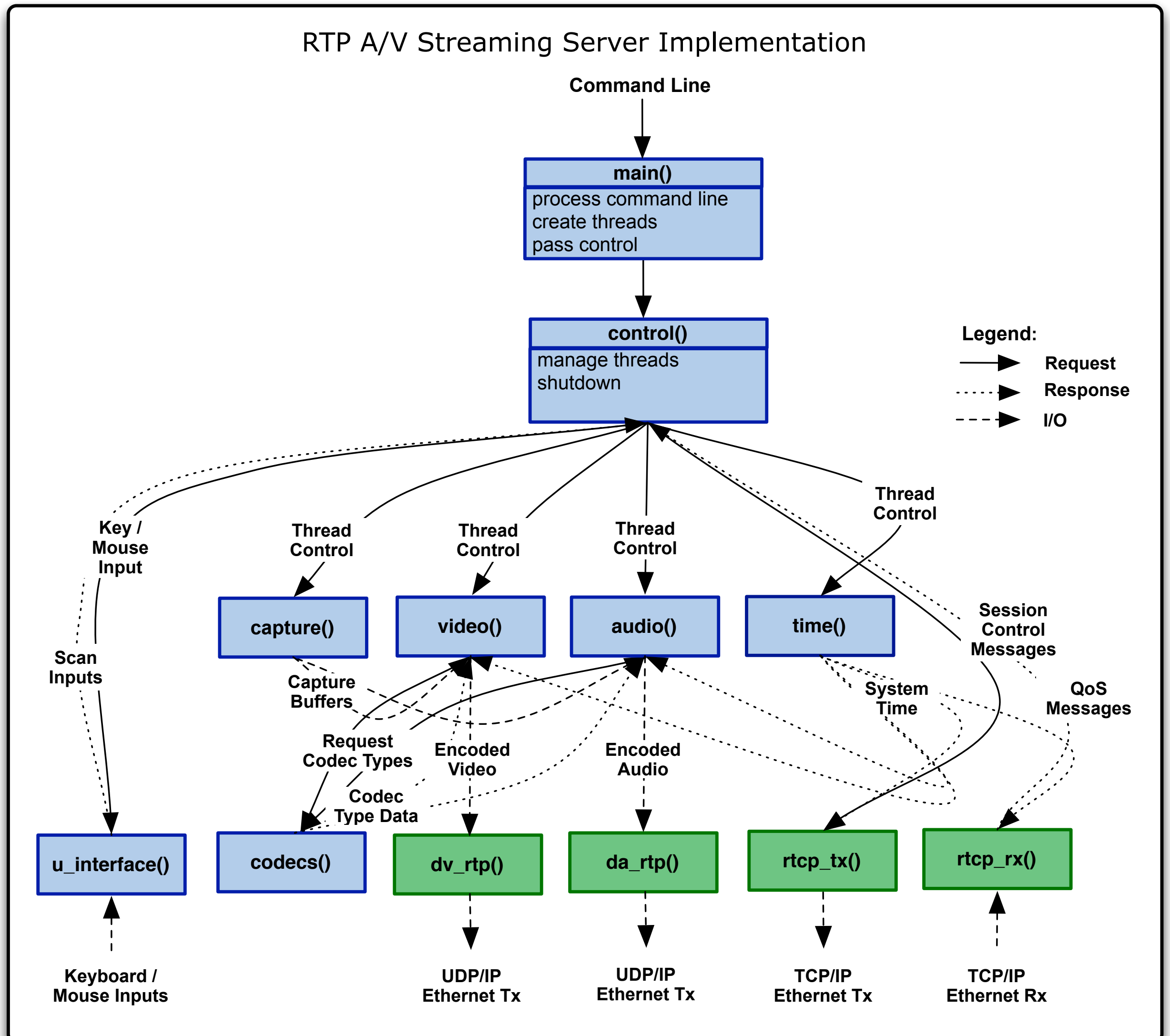

Figure 6: RTP A/V Streaming Client/Server Application Functional Block Diagram.

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 products, please contact us using the contact information below.

Contact Information:

Cimarron Systems, LLC
P.O. Box 1137
Evergreen, Colorado
telephone: (303) 674-9207
email: info@cimarronsystems.com
www.cimarronsystems.com

Revision History:

Date	Version	Notes
11/15/2012	version 1.0	Initial version

