Network Working Group   J. Lazzaro
Request for Comments: 4571   UC Berkeley
Category: Standards Track   July 2006

                Framing Real-time Transport Protocol (RTP)
                  and RTP Control Protocol (RTCP) Packets
                    over Connection-Oriented Transport

Status of This Memo

   This document specifies an Internet standards track protocol for the
   Internet community, and requests discussion and suggestions for
   improvements. Please refer to the current edition of the "Internet
   Official Protocol Standards" (STD 1) for the standardization state
   and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

Abstract

   This memo defines a method for framing Real-time Transport Protocol
   (RTP) and RTP Control Protocol (RTCP) packets onto connection-
   oriented transport (such as TCP). The memo also defines how session
   descriptions may specify RTP streams that use the framing method.

Table of Contents

1. Introduction

   The Audio/Video Profile (AVP, [RFC3550]) for the Real-time Transport
   Protocol (RTP, [RFC3551]) does not define a method for framing RTP
   and RTP Control Protocol (RTCP) packets onto connection-oriented
   transport protocols (such as TCP). However, earlier versions of
   RTP/AVP did define a framing method, and this method is in use in
   several implementations.

   In this memo, we document the framing method that was defined by
   earlier versions of RTP/AVP. In addition, we introduce a mechanism
   for a session description [SDP] to signal the use of the framing
   method. Note that session description signalling for the framing
   method is new and was not defined in earlier versions of RTP/AVP.

1.1. Terminology

   The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
   "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this
   document are to be interpreted as described in BCP 14, RFC 2119
   [RFC2119].

2. The Framing Method

   Figure 1 defines the framing method.

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 ---------------------------------------------------------------
|            LENGTH             |  RTP or RTCP packet ...       |
 ---------------------------------------------------------------
```
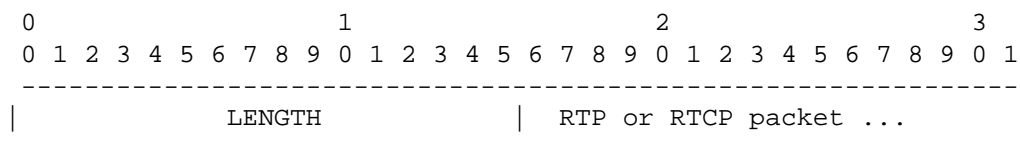
        Figure 1: The bit field definition of the framing method

   A 16-bit unsigned integer LENGTH field, coded in network byte order
   (big-endian), begins the frame. If LENGTH is non-zero, an RTP or
   RTCP packet follows the LENGTH field. The value coded in the LENGTH
   field MUST equal the number of octets in the RTP or RTCP packet.
   Zero is a valid value for LENGTH, and it codes the null packet.

   This framing method does not use frame markers (i.e., an octet of
   constant value that would precede the LENGTH field). Frame markers
   are useful for detecting errors in the LENGTH field. In lieu of a
   frame marker, receivers SHOULD monitor the RTP and RTCP header fields
   whose values are predictable (for example, the RTP version number).
   See Appendix A.1 of [RFC3550] for additional guidance.

3. Packet Stream Properties

   In most respects, the framing method does not specify properties
   above the level of a single packet. In particular, Section 2 does
   not specify the following:

   Bi-directional issues

      Section 2 defines a framing method for use in one direction on a
      connection. The relationship between framed packets flowing in a
      defined direction and in the reverse direction is not specified.

   Packet loss and reordering

      The reliable nature of a connection does not imply that a framed
      RTP stream has a contiguous sequence number ordering. For
      example, if the connection is used to tunnel a UDP stream through
      a network middlebox that only passes TCP, the sequence numbers in
      the framed stream reflect any packet loss or reordering on the UDP
      portion of the end-to-end flow.

   Out-of-band semantics

      Section 2 does not define the RTP or RTCP semantics for closing a
      TCP socket, or of any other "out of band" signal for the
      connection.

   Memos that normatively include the framing method MAY specify these
   properties. For example, Section 4 of this memo specifies these
   properties for RTP/AVP sessions specified in session descriptions.

   In one respect, the framing protocol does indeed specify a property
   above the level of a single packet. If a direction of a connection
   carries RTP packets, the streams carried in this direction MUST
   support the use of multiple synchronization sources (SSRCs) in those
   RTP packets. If a direction of a connection carries RTCP packets,
   the streams carried in this direction MUST support the use of
   multiple SSRCs in those RTCP packets.

4. Session Descriptions for RTP/AVP over TCP

   Session management protocols that use the Session Description
   Protocol [SDP] in conjunction with the Offer/Answer Protocol
   [RFC3264] MUST use the methods described in [COMEDIA] to set up
   RTP/AVP streams over TCP. In this case, the use of Offer/Answer is
   REQUIRED, as the setup methods described in [COMEDIA] rely on
   Offer/Answer.

In principle, [COMEDIA] is capable of setting up RTP sessions for any
RTP profile. In practice, each profile has unique issues that must
be considered when applying [COMEDIA] to set up streams for the
profile.

In this memo, we restrict our focus to the Audio/Video Profile (AVP,
[RFC3551]). Below, we define a token value ("TCP/RTP/AVP") that
signals the use of RTP/AVP in a TCP session. We also define the
operational procedures that a TCP/RTP/AVP stream MUST follow.

We expect that other standards-track memos will appear to support the
use of the framing method with other RTP profiles. The support memo
for a new profile MUST define a token value for the profile, using
the style we used for AVP. Thus, for profile xyz, the token value
MUST be "TCP/RTP/xyz". The memo SHOULD adopt the operational
procedures we define below for AVP, unless these procedures are in
some way incompatible with the profile.

The remainder of this section describes how to setup and use an AVP
stream in a TCP session. Figure 2 shows the syntax of a media (m=)
line [SDP] of a session description:

    "m=" media SP port ["/" integer] SP proto 1*(SP fmt) CRLF

      Figure 2: Syntax for an SDP media (m=) line (from [SDP])

The <proto> token value "TCP/RTP/AVP" specifies an RTP/AVP [RFC3550]
[RFC3551] stream that uses the framing method over TCP.

The <fmt> tokens that follow <proto> MUST be unique unsigned integers
in the range 0 to 127. The <fmt> tokens specify an RTP payload type
associated with the stream.

In all other respects, the session description syntax for the framing
method is identical to [COMEDIA].

The TCP <port> on the media line carries RTP packets. If a media
stream uses RTCP, a second connection carries RTCP packets. The port
for the RTCP connection is chosen using the algorithms defined in
[SDP] or by the mechanism defined in [RFC3605].

The TCP connections MAY carry bi-directional traffic, following the
semantics defined in [COMEDIA]. Both directions of a connection MUST
carry the same type of packets (RTP or RTCP). The packets MUST
exclusively code the RTP or RTCP streams specified on the media
line(s) associated with the connection.

   As noted in [RFC3550], the use of RTP without RTCP is strongly
   discouraged. However, if a sender does not wish to send RTCP packets
   in a media session, the sender MUST add the lines "b=RS:0" AND
   "b=RR:0" to the media description (from [RFC3556]).

   If the session descriptions of the offer AND the answer both contain
   the "b=RS:0" AND "b=RR:0" lines, an RTCP TCP flow for the media
   session MUST NOT be created by either endpoint in the session. In
   all other cases, endpoints MUST establish two TCP connections for an
   RTP/AVP stream, one for RTP and one for RTCP.

   As described in [RFC3264], the use of the "sendonly" or "sendrecv"
   attribute in an offer (or answer) indicates that the offerer (or
   answerer) intends to send RTP packets on the RTP TCP connection. The
   use of the "recvonly" or "sendrecv" attributes in an offer (or
   answer) indicates that the offerer (or answerer) wishes to receive
   RTP packets on the RTP TCP connection.

5. Example

   The session descriptions in Figures 3 and 4 define a TCP RTP/AVP
   session.

   v=0
   o=first 2520644554 2838152170 IN IP4 first.example.net
   s=Example
   t=0 0
   c=IN IP4 192.0.2.105
   m=audio 9 TCP/RTP/AVP 11
   a=setup:active
   a=connection:new

         Figure 3: TCP session description for the first participant

   v=0
   o=second 2520644554 2838152170 IN IP4 second.example.net
   s=Example
   t=0 0
   c=IN IP4 192.0.2.94
   m=audio 16112 TCP/RTP/AVP 10 11
   a=setup:passive
   a=connection:new

         Figure 4: TCP session description for the second participant

   The session descriptions define two parties that participate in a
   connection-oriented RTP/AVP session. The first party (Figure 3) is
   capable of receiving stereo L16 streams (static payload type 11).

The second party (Figure 4) is capable of receiving mono (static
payload type 10) or stereo L16 streams.

The "setup" attribute in Figure 3 specifies that the first party is
"active" and initiates connections, and the "setup" attribute in
Figure 4 specifies that the second party is "passive" and accepts
connections [COMEDIA].

The first party connects to the network address (192.0.2.94) and port
(16112) of the second party. Once the connection is established, it
is used bi-directionally: the first party sends framed RTP packets to
the second party in one direction of the connection, and the second
party sends framed RTP packets to the first party in the other
direction of the connection.

The first party also initiates an RTCP TCP connection to port 16113
(16112 + 1, as defined in [SDP]) of the second party. Once the
connection is established, the first party sends framed RTCP packets
to the second party in one direction of the connection, and the
second party sends framed RTCP packets to the first party in the
other direction of the connection.

6. Congestion Control

The RTP congestion control requirements are defined in [RFC3550]. As
noted in [RFC3550], all transport protocols used on the Internet need
to address congestion control in some way, and RTP is not an
exception.

In addition, the congestion control requirements for the Audio/Video
Profile are defined in [RFC3551]. The basic congestion control
requirement defined in [RFC3551] is that RTP sessions should compete
fairly with TCP flows that share the network. As the framing method
uses TCP, it competes fairly with other TCP flows by definition.

7. Acknowledgements

This memo, in part, documents discussions on the AVT mailing list
about TCP and RTP. Thanks to all of the participants in these
discussions.

8. Security Considerations

Implementors should carefully read the Security Considerations
sections of the RTP [RFC3550] and RTP/AVP [RFC3551] documents, as
most of the issues discussed in these sections directly apply to RTP
streams framed over TCP.

Session descriptions that specify connection-oriented media sessions
(such as the example session shown in Figures 3 and 4 of Section 5)
raise unique security concerns for streaming media. The Security
Considerations section of [COMEDIA] describes these issues in detail.

Below, we discuss security issues that are unique to the framing
method defined in Section 2.

Attackers may send framed packets with large LENGTH values to exploit
security holes in applications. For example, a C implementation may
declare a 1500-byte array as a stack variable, and use LENGTH as the
bound on the loop that reads the framed packet into the array. This
code would work fine for friendly applications that use Etherframe-
sized RTP packets, but may be open to exploit by an attacker. Thus,
an implementation needs to handle packets of any length, from a NULL
packet (LENGTH == 0) to the maximum-length packet holding 64K octets
(LENGTH = 0xFFFF).

9. IANA Considerations

[SDP] defines the syntax of session description media lines. We
reproduce this definition in Figure 2 of Section 4 of this memo. In
Section 4, we define a new token value for the <proto> field of media
lines: "TCP/RTP/AVP". Section 4 specifies the semantics associated
with the <proto> field token, and Section 5 shows an example of its
use in a session description.

10. Normative References

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V.
          Jacobson, "RTP: A Transport Protocol for Real-Time
          Applications", STD 64, RFC 3550, July 2003.

[RFC3551] Schulzrinne, H. and S. Casner, "RTP Profile for Audio and
          Video Conferences with Minimal Control", STD 65, RFC 3551,
          July 2003.

[COMEDIA] Yon, D. and G. Camarillo, "TCP-Based Media Transport in the
          Session Description Protocol (SDP)", RFC 4145, September
          2005.

[SDP]   Handley, M., Jacobson, V., and C. Perkins. "SDP: Session
          Description Protocol", RFC 4566, July 2006.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
          Requirement Levels", BCP 14, RFC 2119, March 1997.

   [RFC3264] Rosenberg, J. and H. Schulzrinne, "An Offer/Answer Model
             with Session Description Protocol (SDP)", RFC 3264, June
             2002.

   [RFC3605] Huitema, C., "Real Time Control Protocol (RTCP) attribute
             in Session Description Protocol (SDP)", RFC 3605, October
             2003.

   [RFC3556] Casner, S., "Session Description Protocol (SDP) Bandwidth
             Modifiers for RTP Control Protocol (RTCP) Bandwidth", RFC
             3556, July 2003.

Author's Address

   John Lazzaro
   UC Berkeley
   CS Division
   315 Soda Hall
   Berkeley CA 94720-1776

   EMail: lazzaro@cs.berkeley.edu

Full Copyright Statement

Intellectual Property

Acknowledgement