John Lazzaro
CS Division
UC Berkeley

Framing RTP and RTCP Packets over Connection-Oriented Transport

<draft-ietf-avt-rtp-framing-contrans-04.txt>

Status of this Memo

Copyright Notice

Abstract

This memo defines a method for framing Real Time Protocol (RTP) and Real Time Control Protocol (RTCP) packets onto connection-oriented transport (such as TCP). The memo also defines how to specify the framing method in a session description.

Table of Contents

1. Introduction 2
1.1 Terminology 2
2. The Framing Method 3
3. Undefined Properties 3
4. Session Descriptions for RTP/AVP over TCP 4
5. Example 5
6. Congestion Control 6
7. Acknowledgements 6
8. Security Considerations 6
9. IANA Considerations 7
10. References 7
10.1 Normative References 7
Authors' Address 8
Intellectual Property Rights Statement 8
Full Copyright Statement 9
Change Log for <draft-ietf-avt-rtp-framing-contrans-04.txt> 10

1. Introduction

The Audio/Video Profile (AVP, [1]) for the Real-Time Protocol (RTP, [2]) does not define a method for framing RTP and Real Time 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 method and show how a session description [4] may specify the use of the method.

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 [5].

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 bitfield 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 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).

3. Undefined Properties

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

The number of RTP or RTCP streams on the connection.

The framing method is commonly used for sending a single RTP or RTCP stream over a connection. However, Section 2 does not define this common use as normative, so that (for example) a memo that defines an RTP SSRC multiplexing protocol may use the framing method.

Bi-directional issues.

Section 2 defines a framing method for use in one direction on a connection. The relationship between framed packets flowing in 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 sessions specified in session descriptions.

4. Session Descriptions for RTP/AVP over TCP

In this section, we show how session descriptions may specify RTP streams that use the framing method.

Figure 2 shows the syntax of a media (m=) line [4] 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 [4]).

The <proto> token value "TCP/RTP/AVP" specifies an RTP/AVP [1] [2] 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 [3].

The TCP <port> on the media line exclusively receives RTP packets. If a media stream uses RTCP, a second connection exclusively receives the RTCP packets. The port for the RTCP connection is chosen using the

algorithms defined in [4] and in related documents.

The TCP connections MAY carry bi-directional traffic, following the semantics defined in [3]. 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.

The RTP stream MUST have an unbroken sequence number order. RTCP stream packets MUST appear as defined in [2], with no lost or re-ordered packets. IETF standards-track documents MAY loosen these restrictions on packet loss and packet ordering.

The out-of-band semantics for the connection MUST comply with [3].

5. Example

The session descriptions in Figure 3-4 define a TCP RTP/AVT 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 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 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** [3].

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 on 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 [4]) of the second party. Once the connection is established, the first party sends framed RTCP packets to the second party on 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 [1]. As noted in [1], 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 [2]. The basic congestion control requirement defined in [2] 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 [1] and RTP/AVP [2] 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 3-4 of Section 5) raise unique security concerns for streaming media. The Security Considerations section of [3] 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.

9. IANA Considerations

[4] 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. References

10.1 Normative References

[1] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson. "RTP: A transport protocol for real-time applications", RFC 3550, July 2003.

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

[3] Yon, D. and G. Camarillo. Connection-Oriented Media Transport in the Session Description Protocol (SDP), draft-ietf-mmusic-sdp-comedia-10.txt.

[4] Handley, M., Jacobson, V., and C. Perkins. "SDP: Session Description Protocol", draft-ietf-mmusic-sdp-new-22.txt.

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

Authors' Address

John Lazzaro
UC Berkeley
CS Division
315 **Soda Hall**
Berkeley CA 94720-1776
Email: lazzaro@cs.berkeley.edu

Intellectual Property Rights Statement

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.

Change Log for <draft-ietf-avt-rtp-framing-contrans-04.txt>

[Note to RFC Editors: this Appendix, and its Table of Contents listing, should be removed from the final version of the memo]

Updated Figures 3 and 4 to use new comedia "connection" attribute.

Changed section numbering to conform to draft-rfc-editor-rfc2223bis-08.txt.