RTP usage with Layered Multimedia Streams

Abstract

This draft describes how one should make use of RTP (rfc1889) when employing layered media streams. This document is meant for implementors of internet multimedia applications that want to use RTP and layered media streams.
1 Introduction

This memorandum describes how to use RTP (Real-Time Transport Protocol, rfc1889) with layered multimedia streams.

2 Layered Video

Today, most multimedia applications place the responsibility of rate-adaptivity at the source. In multicast transmission, source-adaptation, as discussed in [3], leads to the source not being to meet the conflicting bandwidth requirements of all receivers. This usually leads to the least-common denominator scenarios, where the smallest pipe in the network mesh dictates the quality and fidelity of the overall live multimedia "broadcast". If the responsibility of rate-adaptation is placed at the receivers, then heterogeneity of such media transmissions is achievable.

One approach for moving rate-adaptation from the source to the receivers is to combine a layered source-coder with a layered transmission system. In the context of IP Multicast, Deering proposed a realization of this scheme where a source stripes the progressive layers of a hierarchically represented signal across multiple multicast groups [2]. Receivers can then adapt to network heterogeneity by controlling their reception bandwidth through IP Multicast group membership.

In the case of video transmission, several approaches to the layered source-coding problem have been explored, including multirate JPEG [4], subband coding [6], and hierarchical vector quantization [1].

3 RTP Usage

The RTP specification [5] implicitly assumes that the underlying transport/network layer is monolithic. That is, a single RTP session is carried on a single underlying communications layer. However, the layered transmission system described above requires multiple underlying transport end-points. This complicates the naming of an RTP session because we need to explicitly identify each of the transport channels that comprise the overall layered set. When UDP/IP is used for the underlying transport protocol, a session is identified by an IP address A and an even-numbered UDP port number P. RTP data is sent over port P while RTCP control is sent on port P+1 [1,Section 10].

We propose that layered applications use a set of contiguous addresses and ports. Addresses must be distinct because multicast routing and group membership are managed on an address granularity. Ports must be distinct because of a widespread deficiency in existing operating systems, and for unicast, there is only one permissible address. Thus for layer n, the corresponding address is A + n, the data port is P + 2n, and the control port is P + 2n + 1. In the unicast case, only
port mapping applies.

In RTP, each media source is identified with a randomly allocated 32-bit source identifier (SRCID) that is unique only within a single session (a collision detection algorithm resolves conflicts). Additionally, each user is identified with a variable-length "canonical name" (CNAME) string that is globally unique. Data packets are identified only by SRCID, and periodically, each application broadcasts a binding between it's CNAME and SRCID. Thus, a receiver can collate streams across different sessions (identified by different SRCID's) using the level of indirection provided by the CNAME. Using this framework, we can readily handle layered compression formats by treating each layer as a distinct "RTP session" and distributing it on its own multicast group. This is the same approach that the RTP uses to relate separate audio and video streams from a single user.

However, the "RTP session per layer" approach adds unnecessary complexity. Not only does it force each receiver to manage all the CNAME/SRCID bindings, but it requires newly arrived receivers to wait for the binding advertisement before they can start decoding a stream. Another problem is that it creates new error recovery conditions for dealing with conflicting information that arrives on the different sessions.

We propose to extend RTP semantics as follows:

- A single SRCID space is used across all layers and the core (base) layer be used for SRCID allocation and conflict resolution. When a source discovers that it has collided, it transmits an RTCP BYE message on only the base layer.

- A participant sends sender identification (SDES) on only the base layer.

All other RTP rules and practices apply.
4 References


5 Address of the Authors

Michael F. Speer
Sun Microsystems Computer Corporation
2550 Garcia Ave MailStop UMPK14-305
Mountain View, CA 94043
Voice: +1 415 786 6368
Fax: +1 415 786 6445
E-mail: michael.speer@eng.sun.com

Steven McCanne
M/S 50B-2239
Lawrence Berkeley National Laboratory
One Cyclotron Road
Berkeley, CA 94720
Voice: +1 510 486 7520
Fax: +1 510 486 6363
E-mail: mccanne@ee.lbl.gov