ITU Telecommunication Standardization Sector
Study Group 15
Q.2&3/15 Rapporteur Meeting
Eibsee, 24-27 September 1996

Document AVC-1036
September 13, 1996

Source: VideoServer (USA)
Title: Clarification on the use of the RTP G.728 Payload Type
Purpose: Information

References

1 ITU-T Recommendation H.225.0: "Media Stream Packetization and Synchronization of Non -Guaranteed Quality of Service LANs," May 1996.

2 CCITT Recommendation G.728, "Coding of Speech at 16 Kbps Using Low-delay Code Excited Linear Prediction, May 1992.

Motivation

The purpose of this document is to describe the RTP packet format for G.728 audio. In [1], the RTP Payload Type field of 15 designates G728 audio. In G.728 [2], a frame is 2.5 milliseconds of audio and contains four 10 bit code words. Using an example payload of 12 frames (30 milliseconds of audio), the payload content would be 60 bytes. This would be packaged in IP version 4 on Ethernet 2.0 as follows:

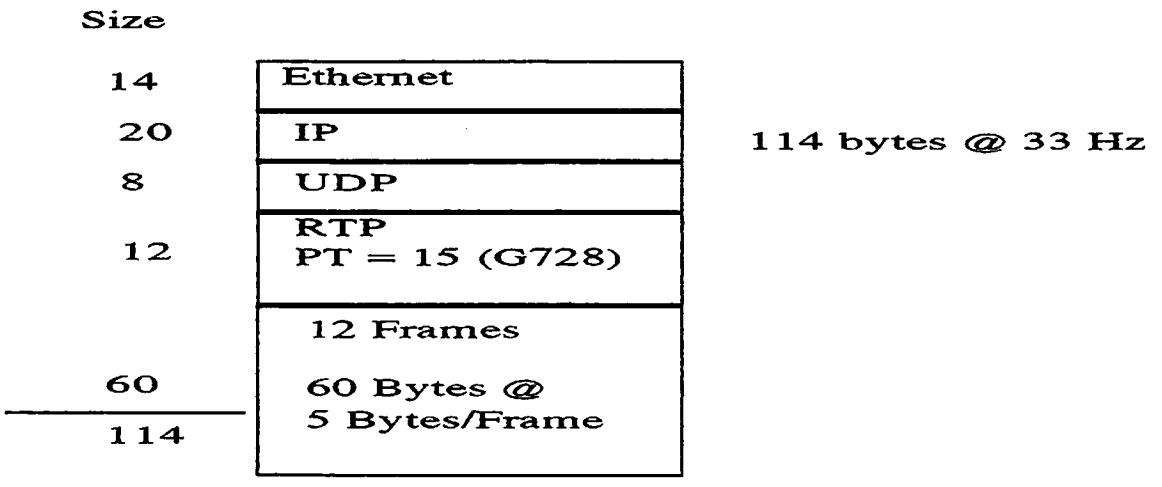

Figure 1: Example H.323 G.728 Audio Packet

The individual frames in the G.728 payload are organized as follows for the first 8 frames:

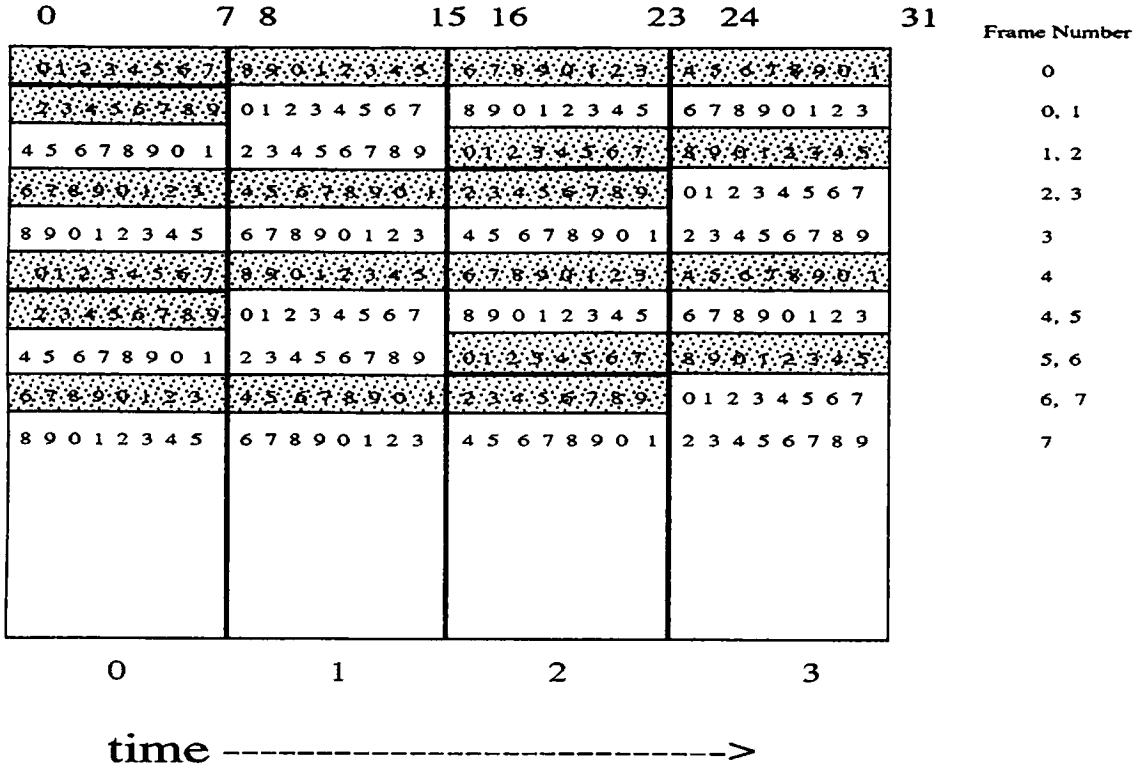

Figure 2: RTP G.728 Frame and Bit Ordering

H.323/H.320 Gateway Considerations

This section discusses some considerations which are important to H.323/H.320 gateway design and concludes with VideoServer recommendations for packet sizes based on our development experience. Figure 1 depicts when G.728 is present in a H.221 multiplex frame.

Octet

	1	2	3	4	5	6	7	8	Bit
0	A	A						FAS	
1	A	A						FAS	
2	A	A						FAS	
*	A	A						*	
15	A	A						FAS	
*	A	A							
*	A	A							
79	A	A							

Figure 3: 10 Millisecond H.221 Frame with G.728

The timing on the audio payload follows:

2	Bits per 125 microseconds
10	Bits per 625 microseconds (a single 10 bit code word)
40	Bits per 2.5 milliseconds (4 code words to a frame)
400	Frames per second

In order to facilitate placing G.728 audio into the H.320 multiplex, it is necessary to start the RTP payload on 40 bit frame boundaries. For the gateway to transcode G.728 into other audio algorithms, it is best to send a discrete number of frames per RTP packet. This allows other packet audio systems to begin play out of the audio without being forced to merge frames from two different RTP packets. Good choices for the number of frames to place in the packet are 12 and 24 frames 30 and 60 milliseconds of audio respectively. These values somewhat simplify transcoding to G.723 audio which contains 30 milliseconds of audio in a frame while still allowing fairly low latency.