

Voice Over IP - Per Call Bandwidth Consumption

Updated: April 13, 2016 **Document ID:** 7934

Contents

Introduction

VoIP - Per Call Bandwidth

- Explanation of Terms

- Bandwidth Calculation Formulas

- Sample Calculation

Configure Voice Payload Sizes in Cisco CallManager and Cisco IOS Gateways

Impact of a Change to Voice Payload Sizes

Voice Activity Detection

RTP Header-Compression or Compressed RTP (cRTP)

- Heuristics for Compression

Related Information

Introduction

This document explains voice codec bandwidth calculations and features to modify or conserve bandwidth when Voice over IP (VoIP) is used. One of the most important factors to consider when you build packet voice networks is proper capacity planning. Within capacity planning, bandwidth calculation is an important factor to consider when you design and troubleshoot packet voice networks for good voice quality.




Note: As a complement to this document, you can use the TAC Voice Bandwidth Codec Calculator (registered customers only) tool. This tool provides information on how to calculate the bandwidth required for packet voice calls.

VoIP - Per Call Bandwidth

These protocol header assumptions are used for the calculations:

- 40 bytes for IP (20 bytes) / User Datagram Protocol (UDP) (8 bytes) / Real-Time Transport Protocol (RTP) (12 bytes) headers.
- Compressed Real-Time Protocol (cRTP) reduces the IP/UDP/RTP headers to 2 or 4 bytes (cRTP is not available over Ethernet).
- 6 bytes for Multilink Point-to-Point Protocol (MP) or Frame Relay Forum (FRF).12 Layer 2 (L2) header.
- 1 byte for the end-of-frame flag on MP and Frame Relay frames.

- 18 bytes for Ethernet L2 headers, which include 4 bytes of Frame Check Sequence (FCS) or Cyclic Redundancy Check (CRC).

 **Note:** This table only contains calculations for the default voice payload sizes in Cisco CallManager or Cisco IOS® Software H.323 gateways. For additional calculations, which includes different voice payload sizes and other protocols, such as Voice over Frame Relay (VoFR) and Voice over ATM (VoATM), use the TAC Voice Bandwidth Codec Calculator (registered customers only) tool.

Codec Information				Bandwidth Calculations					
Codec & Bit Rate (Kbps)	Codec Sample Size (Bytes)	Codec Sample Interval (ms)	Mean Opinion Score (MOS)	Voice Payload Size (Bytes)	Voice Payload Size (ms)	Packets Per Second (PPS)	Bandwidth MP or FRF.12 (Kbps)	Bandwidth w/cRTP MP or FRF.12 (Kbps)	Bandwidth Ethernet (Kbps)
G.711 (64 Kbps)	80 Bytes	10 ms	4.1	160 Bytes	20 ms	50	82.8 Kbps	67.6 Kbps	87.2 Kbps
G.729 (8 Kbps)	10 Bytes	10 ms	3.92	20 Bytes	20 ms	50	26.8 Kbps	11.6 Kbps	31.2 Kbps
G.723.1 (6.3 Kbps)	24 Bytes	30 ms	3.9	24 Bytes	30 ms	33.3	18.9 Kbps	8.8 Kbps	21.9 Kbps
G.723.1 (5.3 Kbps)	20 Bytes	30 ms	3.8	20 Bytes	30 ms	33.3	17.9 Kbps	7.7 Kbps	20.8 Kbps
G.726 (32 Kbps)	20 Bytes	5 ms	3.85	80 Bytes	20 ms	50	50.8 Kbps	35.6 Kbps	55.2 Kbps
G.726 (24 Kbps)	15 Bytes	5 ms			20 ms	50	42.8 Kbps	27.6 Kbps	47.2 Kbps
G.728 (16 Kbps)	10 Bytes	5 ms	3.61	60 Bytes	30 ms	33.3	28.5 Kbps	18.4 Kbps	31.5 Kbps
G722_64k (64 Kbps)	80 Bytes	10 ms	4.13	160 Bytes	20 ms	50	82.8 Kbps	67.6 Kbps	87.2 Kbps

ilbc_m ode_2 0 (15.2K bps)	38 Bytes	20 ms	NA	38 Bytes	20 ms	50	34.0 Kbps	18.8 Kbps	38.4 Kbps
ilbc_m ode_3 0 (13.33 Kbps)	50 Bytes	30 ms	NA	50 Bytes	30 ms	33.3	25.867 Kbps	15.73 Kbps	28.8 Kbps

Explanation of Terms

Code c Bit Rate (Kbps)	Based on the codec, this is the number of bits per second that need to be transmitted in order to deliver a voice call. (codec bit rate = codec sample size / codec sample interval).
Code c Samp le Size (Byte s)	Based on the codec, this is the number of bytes captured by the Digital Signal Processor (DSP) at each codec sample interval. For example, the G.729 coder operates on sample intervals of 10 ms, which corresponds to 10 bytes (80 bits) per sample at a bit rate of 8 Kbps. (codec bit rate = codec sample size / codec sample interval).
Code c Samp le Interv al (ms)	This is the sample interval at which the codec operates. For example, the G.729 coder operates on sample intervals of 10 ms, which corresponds to 10 bytes (80 bits) per sample at a bit rate of 8 Kbps. (codec bit rate = codec sample size / codec sample interval).
Mean Opini on Score (MOS)	MOS is a system used to grade the voice quality of telephone connections. With MOS, a wide range of listeners judge the quality of a voice sample on a scale of one (bad) to five (excellent). The scores are averaged in order to provide the MOS for the codec.
Voice Paylo ad Size	The voice payload size represents the number of bytes (or bits) that are filled into a packet. The voice payload size must be a multiple of the codec sample size. For example, G.729 packets can use 10, 20, 30, 40, 50, or 60 bytes of voice payload size.

(Bytes)	
Voice Payload Size (ms)	The voice payload size can also be represented in terms of the codec samples. For example, a G.729 voice payload size of 20 ms (two 10 ms codec samples) represents a voice payload of 20 bytes [(20 bytes * 8) / (20 ms) = 8 Kbps]
PPS	PPS represents the number of packets that need to be transmitted every second in order to deliver the codec bit rate. For example, for a G.729 call with voice payload size per packet of 20 bytes (160 bits), 50 packets need to be transmitted every second [50 pps = (8 Kbps) / (160 bits per packet)]

Bandwidth Calculation Formulas

These calculations are used:

- Total packet size = (L2 header: MP or FRF.12 or Ethernet) + (IP/UDP/RTP header) + (voice payload size)
- PPS = (codec bit rate) / (voice payload size)
- Bandwidth = total packet size * PPS

Sample Calculation

For example, the required bandwidth for a **G.729** call (8 Kbps codec bit rate) with cRTP, MP, and the default 20 bytes of voice payload is:

- Total packet size (bytes) = (MP header of 6 bytes) + (compressed IP/UDP/RTP header of 2 bytes) + (voice payload of 20 bytes) = 28 bytes
- Total packet size (bits) = (28 bytes) * 8 bits per byte = 224 bits
- PPS = (8 Kbps codec bit rate) / (160 bits) = 50 pps



Note: 160 bits = 20 bytes (default voice payload) * 8 bits per byte

- Bandwidth per call = voice packet size (224 bits) * 50 pps = 11.2 Kbps

Configure Voice Payload Sizes in Cisco CallManager and Cisco IOS Gateways

The voice payload size per packet can be configured in Cisco CallManager and Cisco IOS gateways.



Note: If the Cisco IOS gateway is configured in Cisco CallManager as a Media Gateway Control Protocol (MGCP) gateway, all the codec information (codec type, payload size, voice activity detection, and so on) is controlled by Cisco CallManager.

In Cisco CallManager, the voice payload size per packet is configurable on a systemwide basis. This attribute is set in Cisco CallManager Administration (**Service** > **Service Parameters** > *select_server* > **Cisco CallManager**) with these three service parameters:

- PreferredG711MillisecondPacketSize - (Default setting: 20 ms. Available settings: 10, 20, and 30 ms.)
- PreferredG729MillisecondPacketSize - (Default setting: 20 ms. Available settings: 10, 20, 30, 40, 50, and 60 ms.)
- PreferredG723MillisecondPacketSize - (Default setting: 30 ms. Available settings: 30 and 60 ms.)

In Cisco CallManager, the voice payload size is configured in terms of milliseconds (ms) samples. Based on the codec, this table maps some ms samples to the actual payload size in bytes.

C o d e c	Voice Payload Size (ms)	Voice Payload Size (Bytes)	Comments
G. 7 1 1	20 ms (default)	160 Bytes	Notice that the codec bit rate is always maintained. For example: G.711 codec = [240 bytes * 8(bits/bytes)] / 30 ms = 64 Kbps
	30 ms	240 Bytes	
G. 7 2 9	20 ms (default)	20 Bytes	
	30 ms	30 Bytes	
G. 7 2 3	30 ms (default)		

In Cisco IOS gateways, a feature is added in Cisco IOS Software Release 12.0(5)T that allows the voice payload size (in bytes) for VoIP packets to be changed through the CLI. The new command syntax follows:

```
Cisco-Router(config-dial-peer)#codec g729r8 bytes ?
```

Each codec sample produces 10 bytes of voice payload.

Valid sizes are:

```
10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120,
130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230
```

Any other value within the range will be rounded down to nearest valid size.

```
<10-230> Choose a voice payload size from the list above
```

Impact of a Change to Voice Payload Sizes

The number of codec samples per packet is another factor that determines the bandwidth and delay of a VoIP call. The codec defines the size of the sample, but the total number of samples placed in a packet affects how many packets are sent per second.

When you increase the voice payload size the VoIP bandwidth reduces and the overall delay increases. This example illustrates this:

- **G.729** call with voice payload size of 20 bytes (20 ms): (40 bytes of IP/UDP/RTP headers + 20 bytes voice payload) * 8 bits per byte * 50 pps = 24 Kbps
- **G.729** call with voice payload size of 40 bytes (40 ms): (40 bytes of IP/UDP/RTP headers + 40 bytes voice payload) * 8 bits per byte * 25 pps = 16 Kbps

**Notes:**

- L2 headers are not considered in this calculation.
- The calculations show that while the payload size is doubled, the number of packets per second required is subsequently cut in half.
- As defined in the International Telecommunication Union Telecommunication Standardization Sector (ITU-T) G.114 specifications, the recommended one-way overall delay for voice is 150 ms. For a private network, 200 ms is a reasonable goal, and 250 ms must be the maximum.

Voice Activity Detection

With circuit-switched voice networks, all voice calls use 64 Kbps fixed-bandwidth links regardless of how much of the conversation is speech and how much is silence. With VoIP networks, all conversation and silence is packetized. With Voice Activity Detection (VAD), packets of silence can be suppressed.

Over time and as an average on a volume of more than 24 calls, VAD can provide up to a 35 percent bandwidth savings. The savings are not realized on every individual voice call, or on any specific point measurement. For the purposes of network design and bandwidth engineering, VAD must not be taken into account, especially on links that carry fewer than 24 voice calls simultaneously. Various features such as music on hold and fax render VAD ineffective. When the network is engineered for the full voice call bandwidth, all savings provided by VAD are available to data applications.

VAD also provides Comfort Noise Generation (CNG). Because you can mistake silence for a disconnected call, CNG provides locally generated white noise so the call appears normally connected to both parties. **G.729** Annex-B and G.723.1 Annex-A include an integrated VAD function, but otherwise performs the same as **G.729** and G.723.1, respectively.

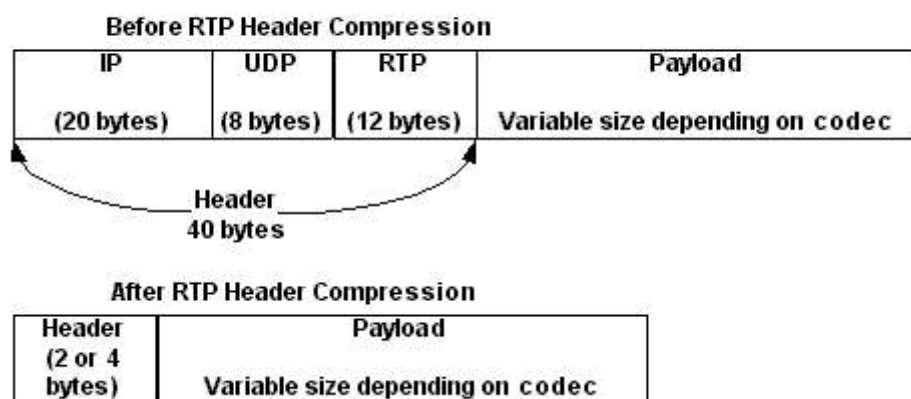
In Cisco CallManager, VAD can be enabled (it is disabled by default) with these service parameters:

- **SilenceSuppressionSystemWide** - This parameter selects the VAD setting for all skinny endpoints (for example, Cisco IP Phones and Skinny gateways)
- **SilenceSuppressionWithGateways** - This parameter selects the VAD setting for all MGCP gateways. This does not have an effect on H.323 gateways. VAD on H.323 gateways must be disabled on the gateway.

You can find these service parameters under Cisco CallManager Administration (**Service > Service Parameters > select_server > Cisco CallManager**).

RTP Header-Compression or Compressed RTP (cRTP)

RTP Header Compression



All VoIP packets are made up of two components: voice samples and IP/UDP/RTP headers. Although the voice samples are compressed by the Digital Signal Processor (DSP) and can vary in size based on the codec used, these headers are a constant 40 bytes in length. When compared to the 20 bytes of voice samples in a default **G.729** call, these headers make up a considerable amount of overhead. With cRTP, these headers can be compressed to two or four bytes. This compression offers significant VoIP bandwidth savings. For example, a default **G.729** VoIP call consumes 24 Kb without cRTP, but only 12 Kb with cRTP enabled.

Because cRTP compresses VoIP calls on a link-by-link basis, both ends of the IP link need to be configured for cRTP.

In Cisco IOS Software Releases 12.0.5T and earlier, cRTP is process-switched, which severely limits the scalability of cRTP solutions due to CPU performance. Most of these issues have been resolved through various cRTP performance improvements introduced in Cisco IOS Software Releases 12.0.7T through 12.1.2T. This is a summary of the history.

- cRTP is process-switched in Cisco IOS Software Release 12.0.5T and earlier.
- In Cisco IOS Software Release 12.0.7T, and continuing in 12.1.1T, fast-switching and Cisco Express Forwarding-switching support for cRTP is introduced.
- In Cisco IOS Software Release 12.1.2T, algorithmic performance improvements are introduced.

Moving cRTP into the fast-switching path significantly increases the number of RTP sessions (VoIP calls) that VoIP gateways and intermediate routers can process.


Heuristics for Compression

As RTP does not have a distinct packet header of its own, an RTP stream (for cRTP) is distinguished from a UDP stream (cUDP) by the use of heuristics. The exact heuristics used at present in order to detect RTP packets for compression are:

- The destination port number is even.
- The destination port number is in the range 16384–32767 or 49152–65535.
- The RTP version field is set to two.
- The RTP extension field is set to zero.

Related Information

- Voice Technology Support
- Voice and Unified Communications Product Support

- Troubleshooting Cisco IP Telephony  leavingcisco.com
- Technical Support - Cisco Systems

© 2019 Cisco and/or its affiliates. All rights reserved.