Back to the Basics:

Audio Codec Selection, Configuration, and its Impact on the Network

Toby Sauer
Instructor/Consultant
November 8, 2023

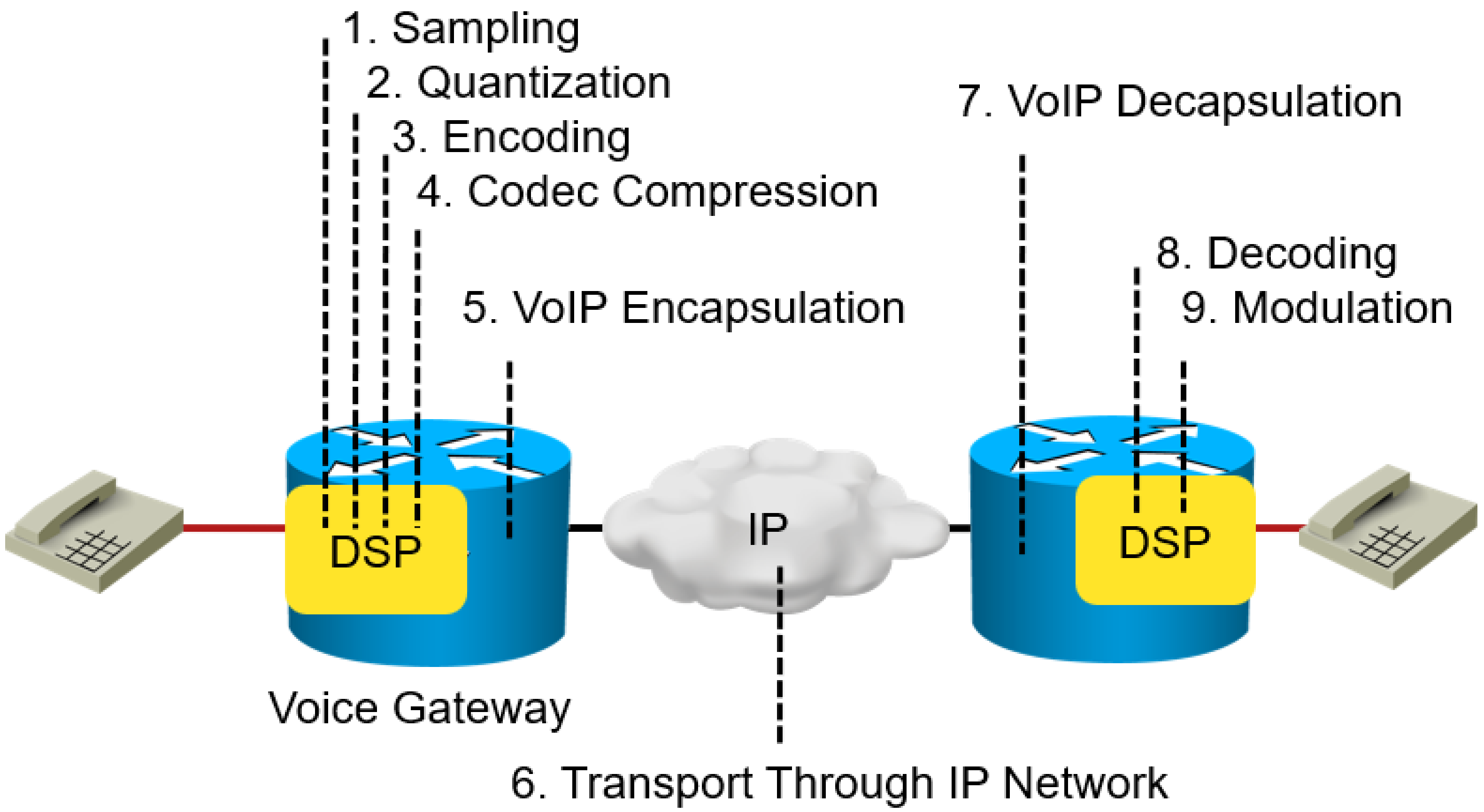
Our Agenda

- Voice Packetization
- VoIP Media Transmission
- Voice Activity Detection
- Evaluating Quality of Codecs
- Evaluating Overhead
- Codec Selection in CUCM
- Codec Selection in IOS Devices
- Closing & Wrap Up
  Your Questions and Feedback

athena

Voice Packetization

Major Stages of Voice Processing in VoIP



Converting Voice to VoIP

Overview

Pulse Code Modulation (PCM)

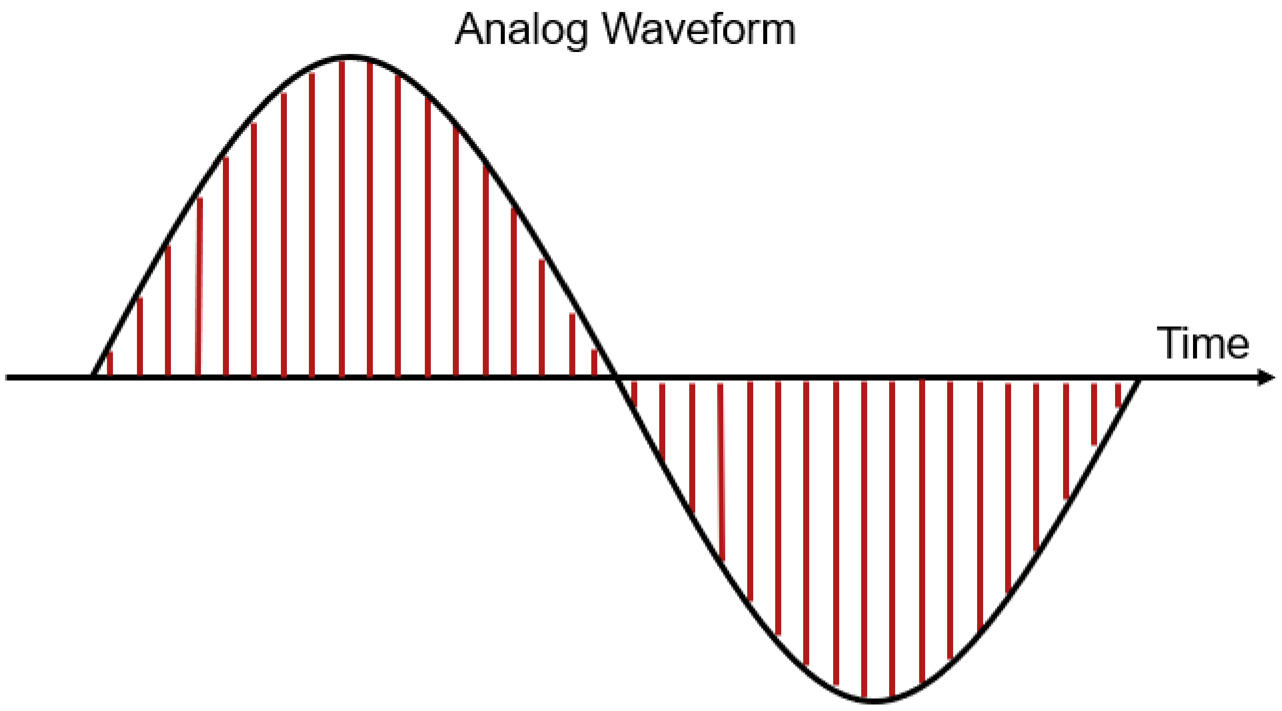
1. Sample the analog signal regularly.
2. Quantize the sample.
3. Encode the value into a binary expression.

Codec Compression

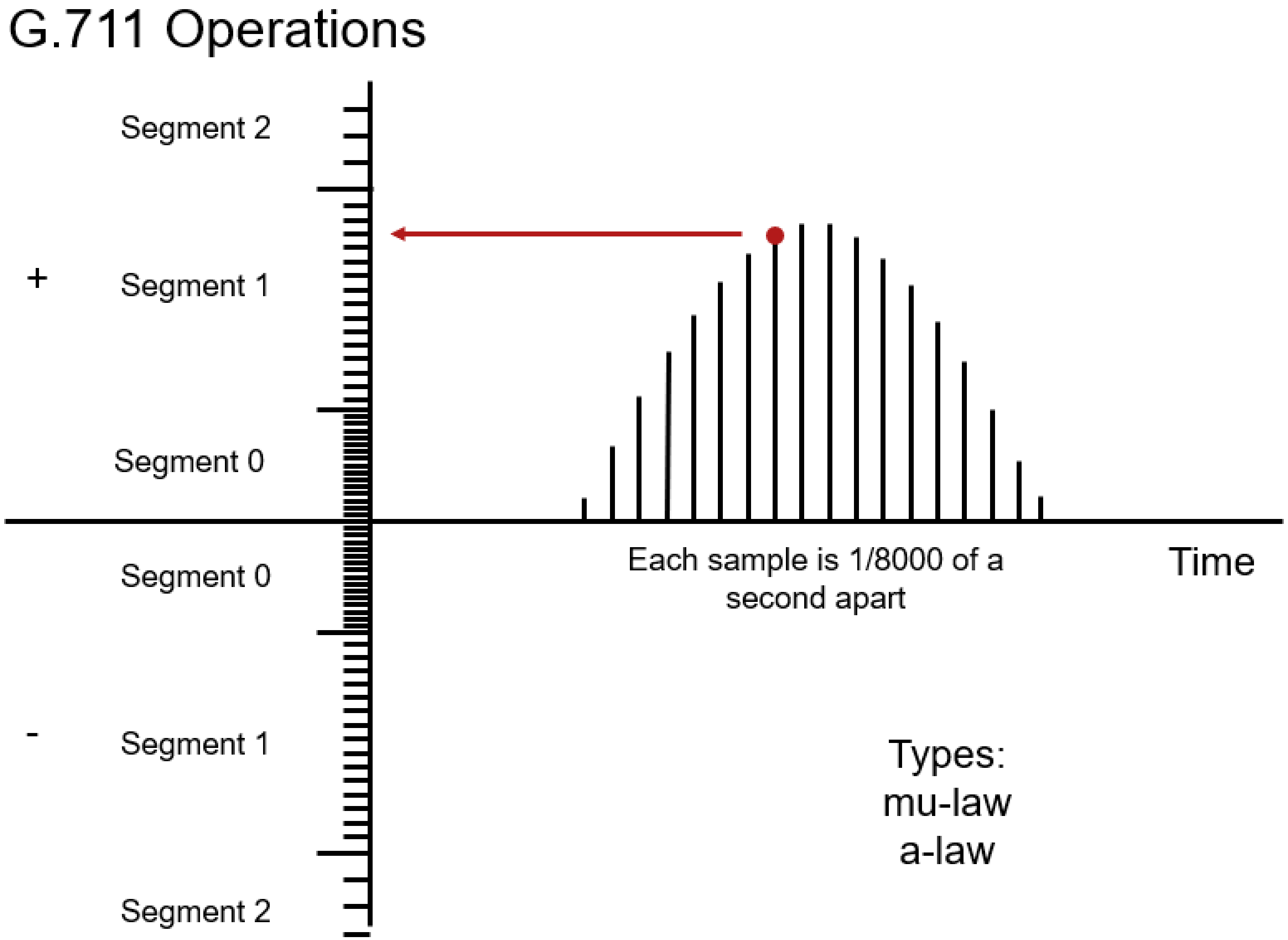
4. Compress the samples to reduce bandwidth (optional).

Sampling

- Significant human articulation range:
  - 300 Hz to 4 kHz
- Nyquist theorem: sampling rate = 2 x maximum articulation frequency
  - 2 x 4 kHz = 8 kHz = 8000/sec
  - Each sample is 1/8000 of a second apart



Quantization



Quantization (Cont.)

G.711 a-law and mu-law

Similarities between mu-law and a-law
Both are linear approximations of logarithmic input/output relationship.
Both are implemented using eight-bit codewords (256 levels, one for each quantization interval).
Both break the range into a total of 16 segments: • Eight positive and eight negative segments. • Each segment is twice the length of the preceding one. • Uniform quantization in each segment.
Both use a similar approach to coding the eight-bit word: • First (MSB) identifies polarity. • Bits two, three, and four identify segment. • Final four bits quantize the segment are the lower signal levels than a-law.

Differences between mu-law and a-law
Different linear approximations lead to different lengths and slopes.
The numerical assignment of the bit positions in the eight-bit codeword to segments and the quantization levels within segments are different.
a-law provides a greater dynamic range than u-law.
mu-law provides better signal/distortion performance for low-level signals than a-law.
• An international connection needs to use a-law • mu to a conversion is the responsibility of the mu-law country.

Encoding

G.711 8-Bit Words

1	0	0	1	1	1	0	0

Segment

Interval

Sign:

1 = Positive

0 = Negative

Example: mu-law = +99 and a-law = +28

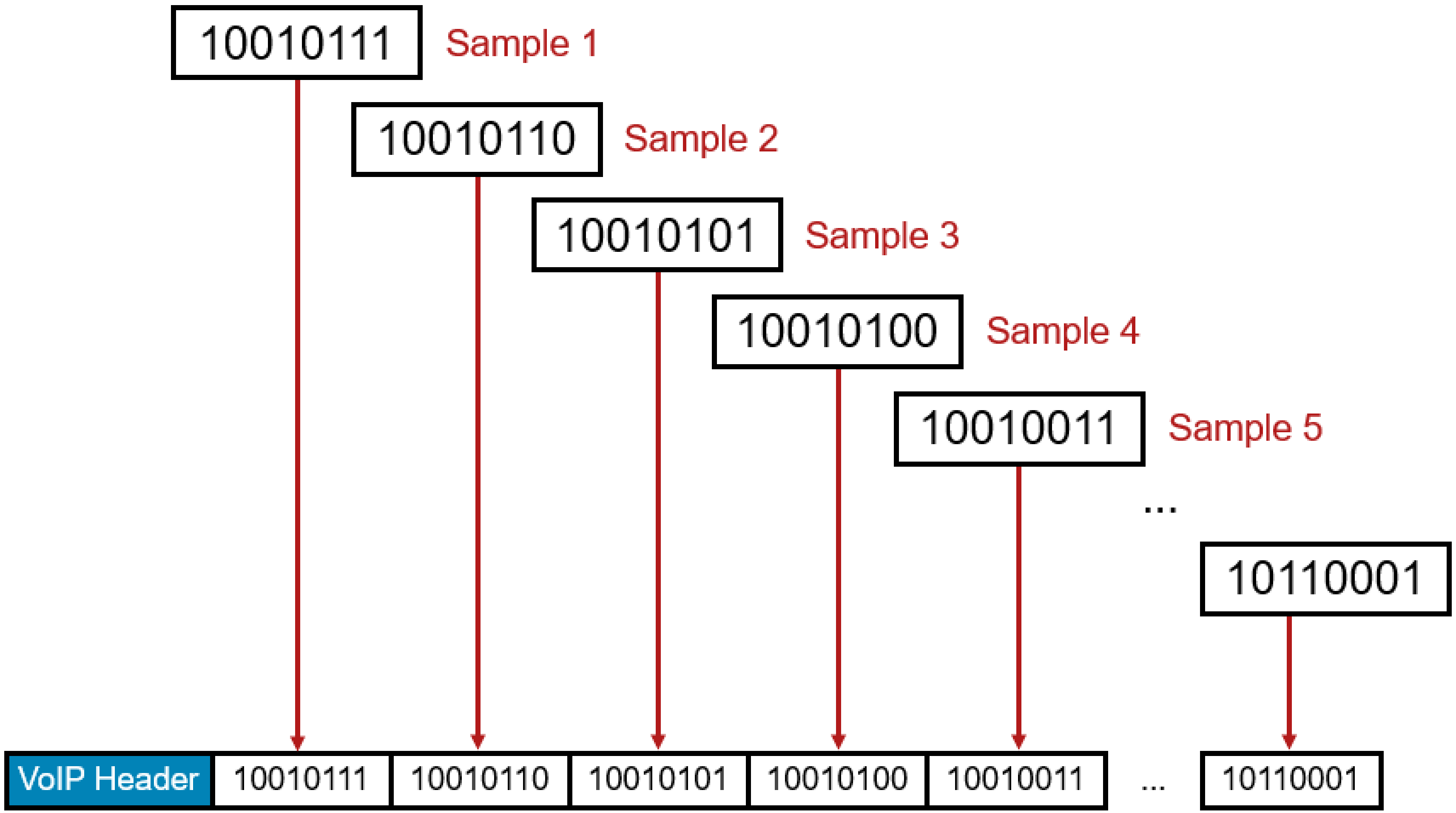
Compression

Codec	Bandwidth (kbps)	Developed	Band	Sampling (kHz)	MOS
G.711	64	ITU-T	Narrowband	8	4,3
G.722	48 / 56 / 64	ITU-T	Wideband	16	---
iLBC	15,2 / 13,3	Global IP Solutions*	Narrowband	8/16	4,14
OPUS	6-510	IETF	Mixed	8/12/16/24/48	---
G.729	8	ITU-T	Narrowband	8	3,92

* Acquired by Google in 2011.

** Since January 2017.

PCM (G.711)



G.711 20 ms of samples (160 bytes)
G.711 30 ms of samples (240 bytes)

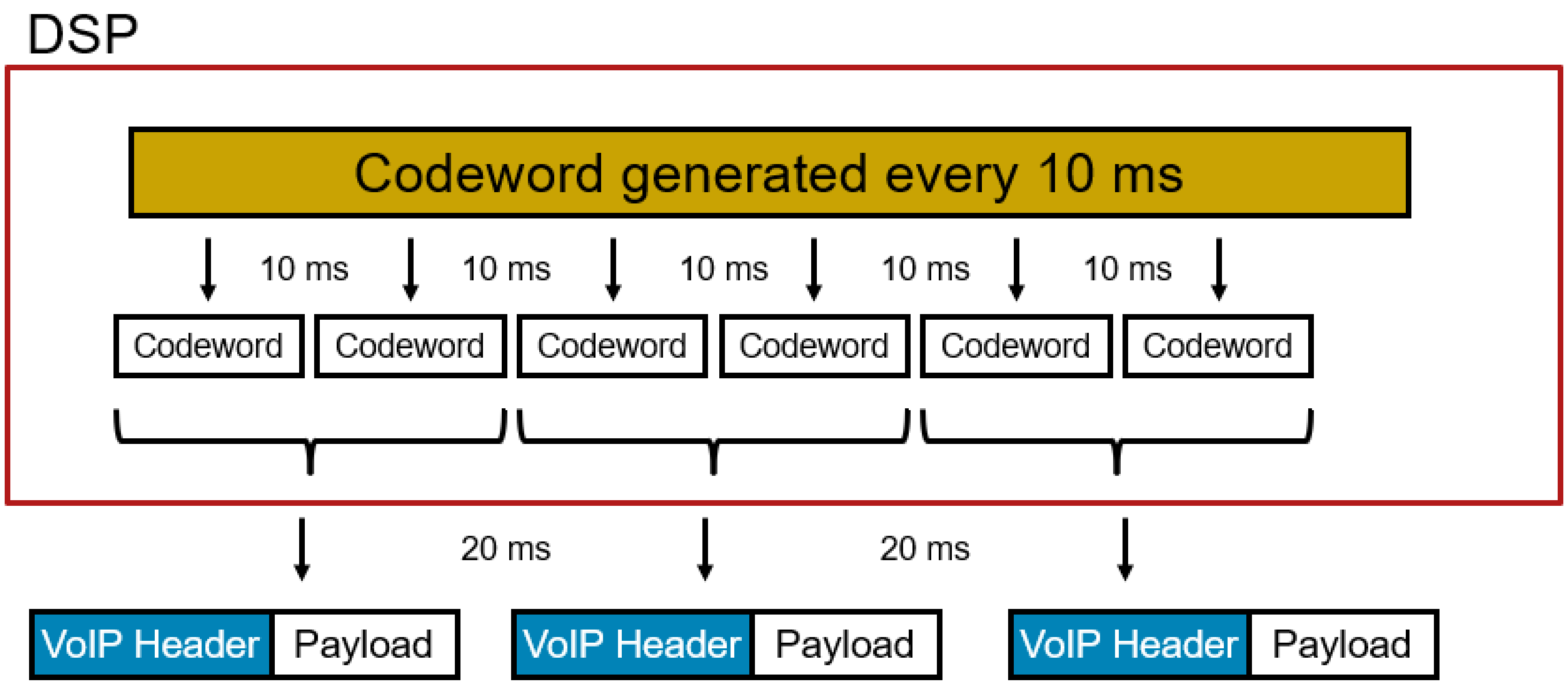
Packetization Rate

- The length of voice streams in a packet affects packetization rate, sample size, and voice bandwidth.

	20-ms voice length in a packet	30-ms voice length in a packet	40-ms voice length in a packet	60-ms voice length in a packet	80-ms voice length in a packet
Packetization rate	50 p/s	33.3 p/s	25 p/s	16.7 p/s	12.5 p/s
Size of collected G.711 samples for a single packet	160 B	240 B	320 B	480 B	640 B
Uncompressed raw voice bandwidth	64 kb/s	64 kb/s	64 kb/s	64 kb/s	64 kb/s
Layer 3+ uncompressed VoIP bandwidth	80 kb/s	74.7 kb/s	72 kb/s	69.3 kb/s	68 kb/s

Codec Operations

G.729



By default, one packet contains 20 ms of voice: 2 codewords
30 ms packetization period: 3 codewords in one packet

Packetization and Compression Example

G.729

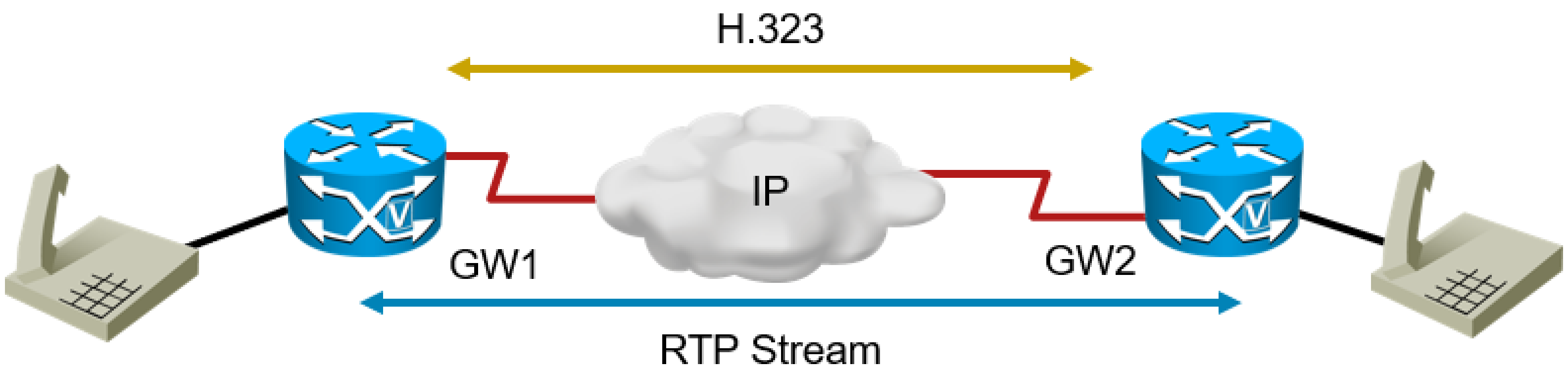
- Layer 3 + bandwidth per call = (Voice Payload + Layer 3 Overhead [40B]) x Packets per Second x 8 bits/Byte

	20 ms voice length in a packet	30ms voice length in a packet
Packetization rate	50 p/s	33.3 p/s
Size of collected, compressed G.729 samples for a single packet	20 B	30 B
Compressed raw voice bandwidth	8 kb/s	8 kb/s
Layer 3+ G.729 VoIP bandwidth	24 kb/s	18.7 kb/s

VoIP Media Transmission

VoIP Media Transmission Overview

- Real-Time Transport Protocol: Delivers the actual audio and video streams over networks.
- Real-Time Transport Control Protocol: Provides out-of-band control information for an RTP flow.
- cRTP compresses IP/UDP/RTP headers on low-speed serial links.
- SRTP provides encryption, message authentication and integrity, and replay protection to the RTP.



Real-Time Transport Protocol

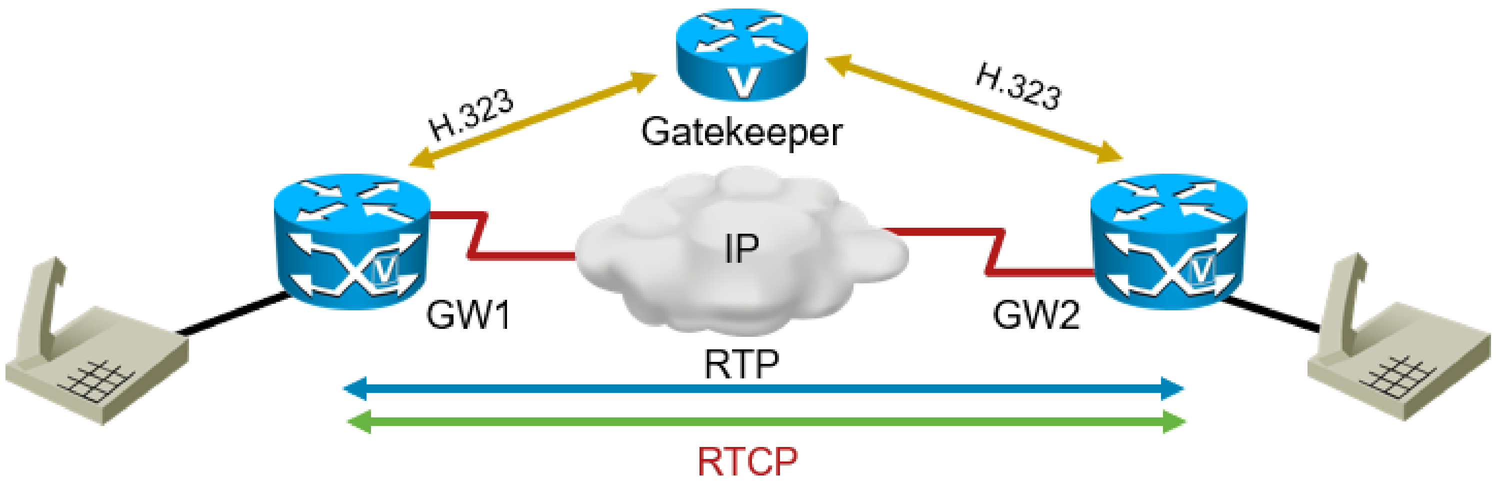
- Provides end-to-end delivery for real-time data, such as voice and video
- Randomly picks even ports from UDP port range 16384–32767
- Includes the following services:
  - Payload type identification (codec type and media format)
    - Allows the codec to change during transmission, as with fax/modem pass-through
  - Sequence numbering
    - Primarily to detect packet loss
  - Measuring delay/jitter
    - To place packets in the correct timing order (playout delay compensation)

Layer 2 Header	IP Header	UDP Header	RTP Header	Voice Payload

Flags	Payload type	Sequence number	Time stamp	Options

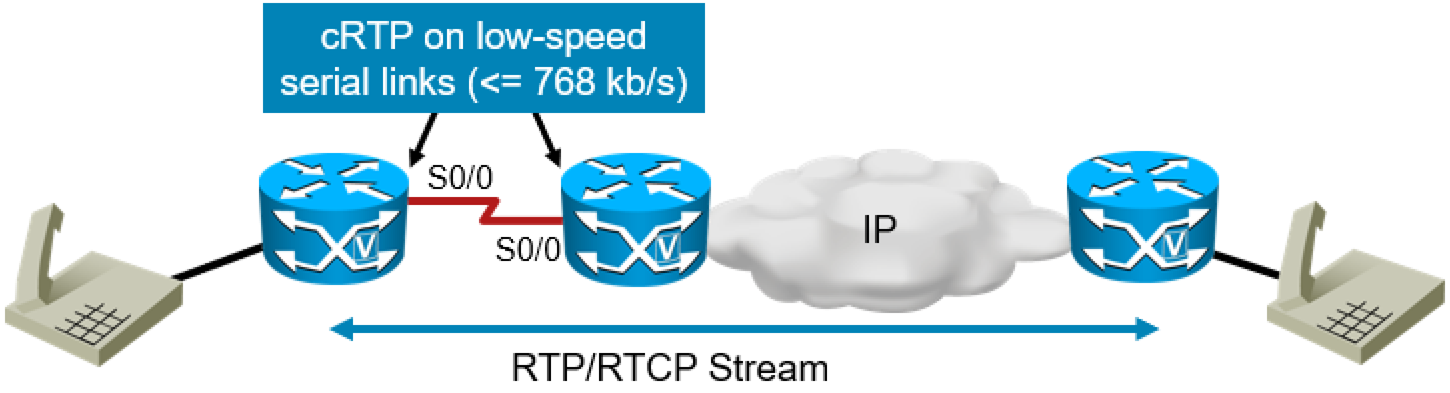
Real-Time Transport Control Protocol

- Monitors media quality and provides control information
- Provides feedback on the RTP session:
  - Packet count
  - Packet delay
  - Octet count
  - Packet loss
  - Jitter (variation in delay)
- Provides a separate flow from RTP for UDP transport use:
  - Is paired with its RTP stream
  - RTP stream UDP port plus 1 (odd-numbered port)



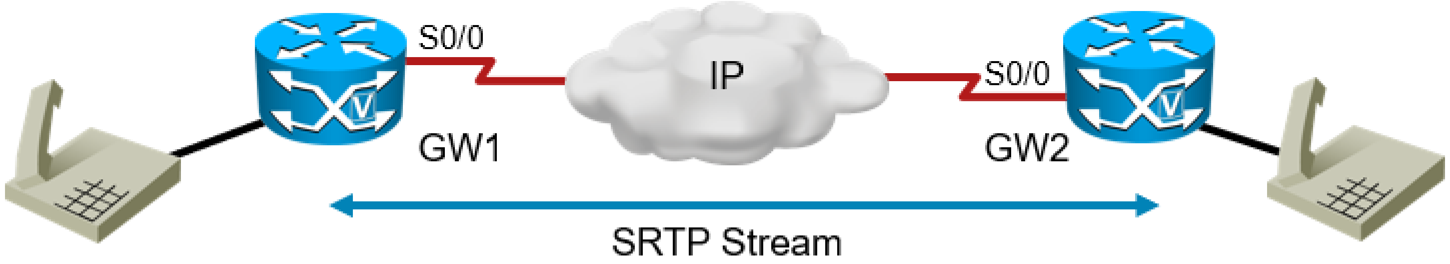
Compressed RTP

- Maps 40-byte header to 2 (without checksum) or 4 (with checksum) bytes most of the time
- Works point-to-point, must be configured on both ends of the link
- On high-speed links, processing overhead does not justify the bandwidth savings
- Algorithm:
  - Establishes session context (full IP/UDP/RTP headers, few first-order differential values, link sequence number, generation number, and a delta encoding table)
  - Session state shared between the compressor and the decompressor
  - After the context state is established, compressed packets may be sent
  - Only change (delta) indicators are transmitted

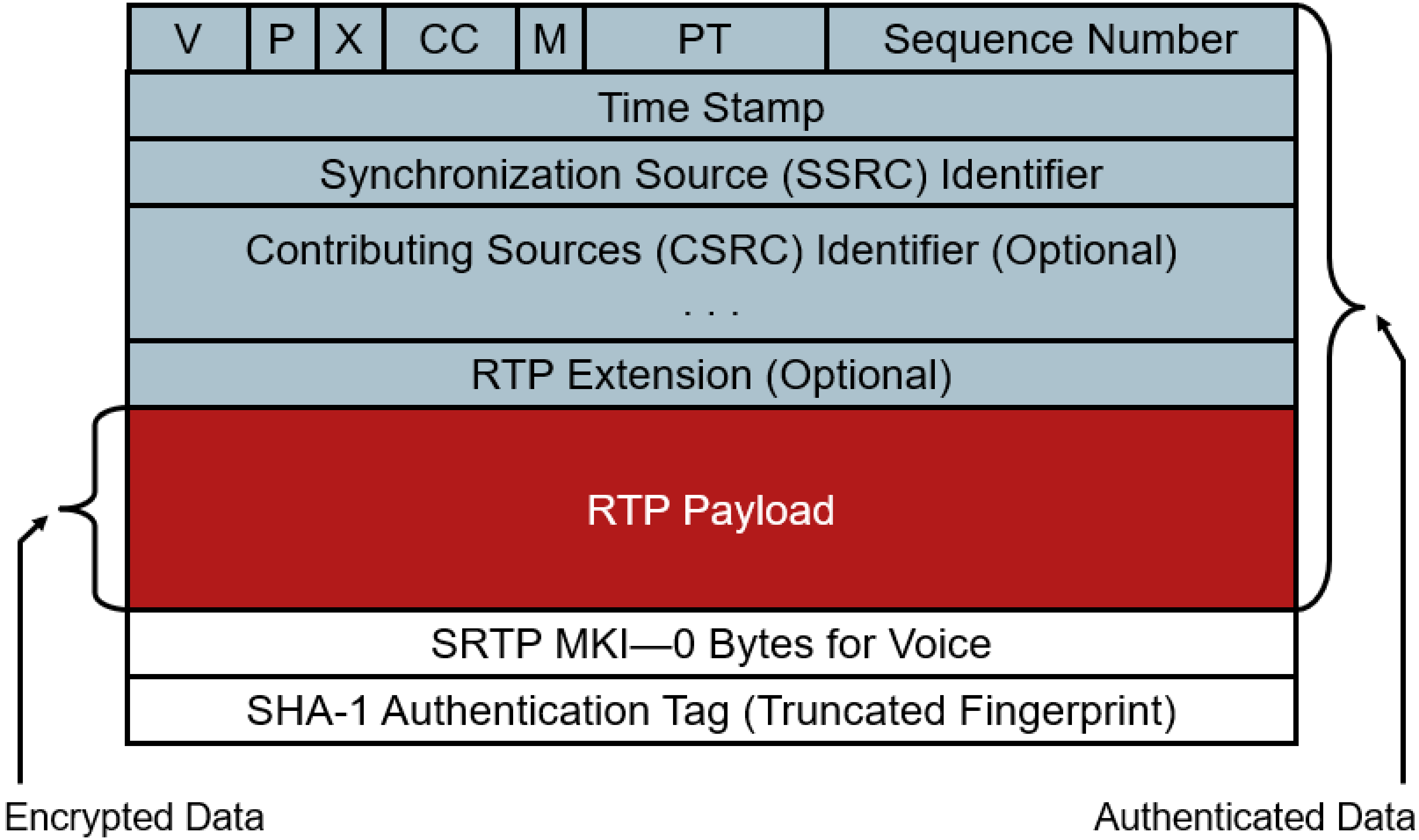


Secure RTP

- Encryption
  - Makes the content undecipherable for transit
- Message integrity
  - Adds a fingerprint to detect tampering during transit
- Message authentication
  - Protects the fingerprint with key to guarantee the authenticity of the source
- Replay protection
  - Sequencing prevents the injecting of outdated information



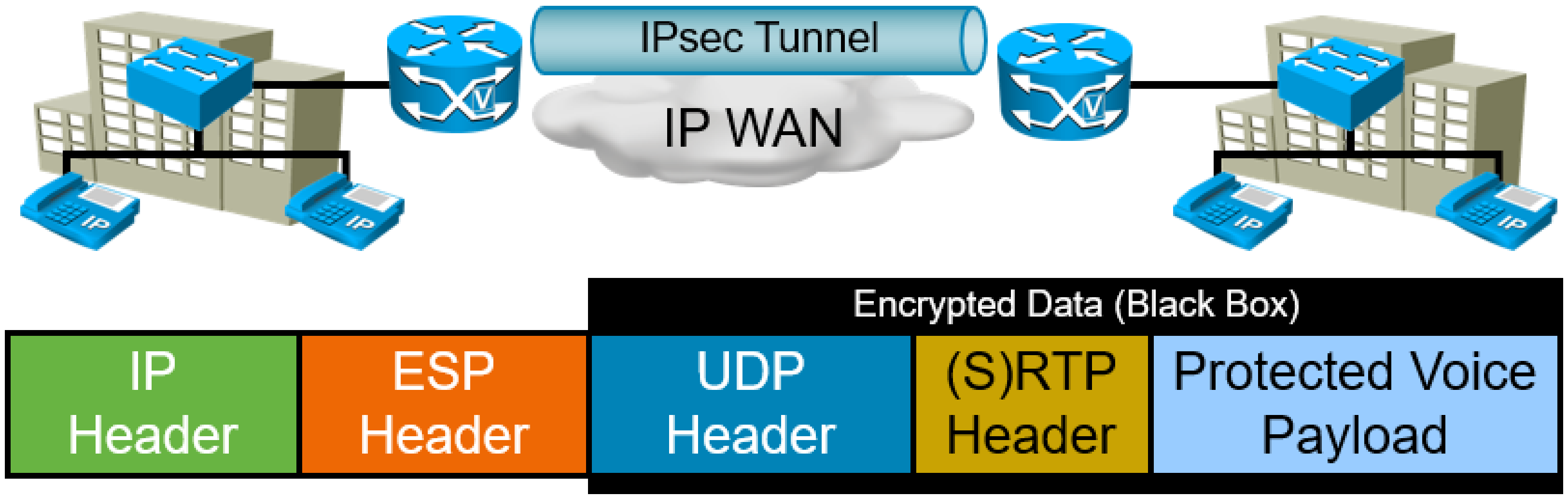
Secure RTP Packet Format



VoIP Media Considerations

Privacy

- IPsec protection of SRTP packets encrypts already-encrypted data
- Exclude SRTP packets from IPsec protection:
  - To save bandwidth and computational resources
- Prefer SRTP over IPsec:
  - Less overhead
  - More uniform approach (covers other calls, such as from roaming users)

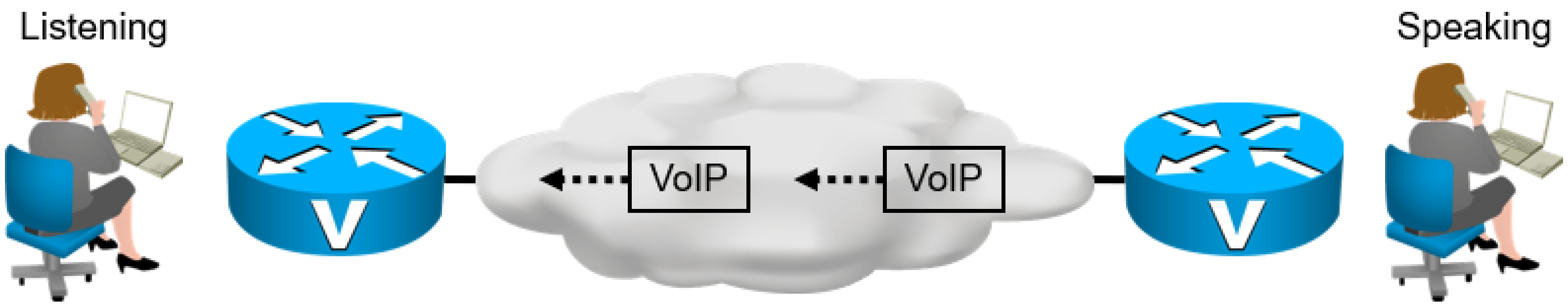


athena

Voice Activity Detection

Voice Activity Detection Overview

- Builds on the nature of human conversation
  - One speaks, one listens
- Classifies packets into: speech, silence, and unknown
  - Speech and unknown packets are sent over the network
  - Packets that would carry silence are discarded
- Up to 35 percent bandwidth savings
  - Based on average volume of more than 24 calls
- The sound quality could be slightly degraded by VAD
  - Initial after-silence sounds chopped off



Bandwidth Savings

Codec	Codec speed	Sample size	Frame Relay, no VAD	Frame Relay with VAD
G.711	64 kb/s	240 B	76.3 kb/s	49.6 kb/s
G.711	64 kb/s	160 B	82.4 kb/s	53.6 kb/s
iLBC	13.3 kb/s	30 B	26.1 kb/s	17.0 kb/s
iLBC	15.2 kb/s	20 B	34.4 kb/s	22.4 kb/s
G.729	8 kb/s	30 B	20.3 kb/s	13.2 kb/s
G.729	8 kb/s	20 B	26.4 kb/s	17.2 kb/s

athena

Evaluating Quality of Codecs

Skyline Advanced Technology Services

Voice Quality Evaluation

Test Methods

Mean opinion score (MOS):

- Defined in ITU-T Recommendation P.800
- Results in subjective measures
- Scores from 1 (worst) to 5 (best); 4.0 is business quality

Perceptual Evaluation of Speech Quality (PESQ):

- Automated assessment of the speech quality as experienced by users
- Successor of Perceptual Speech Quality Measurement (PSQM)
- ITU-T recommendation P.862 (Feb 2001)
- Worldwide applied industry standard for objective voice quality testing
- PESQ results principally model mean opinion scores (MOSs)

Perceptual Evaluation of Audio Quality (PEAQ):

- Automated assessment of speech and other audio types
- Patented and available under license
- PEAQ results principally model MOSs

Voice Quality Evaluation

Test Methods Comparison

Feature	MOS	PSQM	PESQ	PEAQ
Test method	Subjective	Objective	Objective	Objective
End-to-end packet loss test	Inconsistent	No	Yes	Yes
End-to-end jitter test	Inconsistent	No	Yes	Yes
Measurement subject	Voice and other audio	Voice	Voice	Voice and other audio

Codec Quality

Codec	Bandwidth [kb/s]	MOS *
G.711	64	4.3
G.726r32	32	3.8
G.726r24	24	3.75
G.726r16	16	3.7
G.728	16	3.75
iLBC	15.2	4.14
GSM Full Rate	13	3.5
G.729	8	3.92
G.729a	8	3.7
G.723r63	6.3	3.7
G.723r53	5.3	3.65

*MOS values under ideal network conditions: no packet loss, low delay, and no jitter

Evaluating Overhead

Evaluating Overhead

Bandwidth Calculation

- The table lists Layer 3+ bandwidth per call, excluding overhead.
- More accurate bandwidth per call calculation would include Layer 2 overhead.

Codec	Packetization Period	Voice Payload	Packets per Second	Layer 3+ Bandwidth per Call
G.711	20 ms	160 Byte	50	80 kb/s
G.711	30 ms	240 Byte	33	74 kb/s
G.729	20 ms	20 Byte	50	24 kb/s
G.729	30 ms	30 Byte	33	19 kb/s

Bandwidth per call =
(Voice payload + Layer 3+ overhead + Layer 2 overhead) * packets per second * 8 bits/byte

Evaluating Overhead (Cont.)

Layer 2 and Layer 3+ Overhead

Layer 2 Headers [Bytes]	
802.3 Ethernet	18
802.1Q Ethernet	18+4
PPP	6-9
Multilink PPP with Interleaving	13
Frame Relay	6
Frame Relay with FRF.12	8

Layer 3 + Headers [Bytes]	
IP	20
UDP	8
RTP	12

VPN Headers [Bytes]	
ESP	50-57
GRE/L2TP	24
MPLS label	4

Evaluating Overhead (Cont.)

Bandwidth Calculation Example

- Example: Layer 3+, G.711 over Frame Relay, 50 Packets per Second

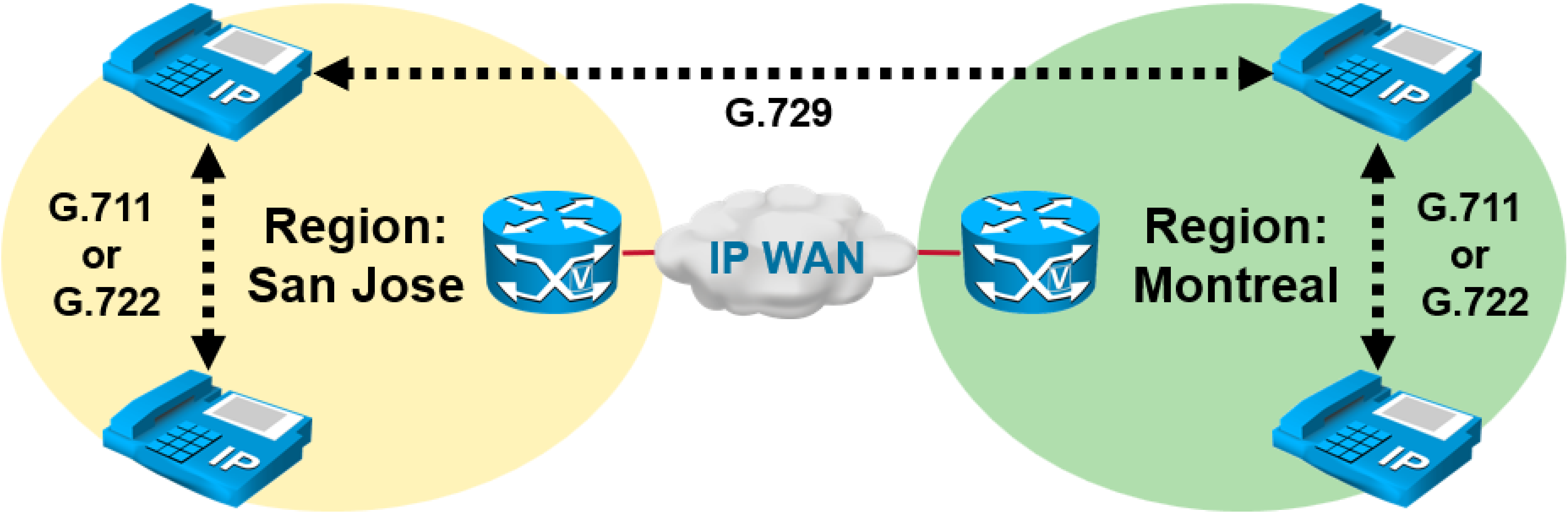
Bandwidth per call
= (Voice payload + Layer 3 OH + Layer 2 OH) * packets per second x 8 bits/byte
= (160 + 40 + 6) bytes * 50 pps * 8 bit/byte
= 82,400 b/s = 82.4 kb/s

Per-Call Bandwidth Using Common Codecs

Codec	Voice Payload	PPS	Only Layer 3+	Call over Frame Relay	Call over 802.3 Ethernet
G.711	160 bytes	50	80 kb/s	82.4 kb/s	87.2 kb/s
G.711	240 bytes	33	74.66 kb/s	76.27 kb/s	79.47 kb/s
G.729	20 bytes	50	24 kb/s	26.4 kb/s	31.2 kb/s
G.729	30 bytes	33	18.66 kb/s	20.27 kb/s	23.47 kb/s

Codec Selection in CUCM

- G.711 / G.722 codecs uses the most bandwidth (64kbps per call) and is typically used within the LAN.
- G.729 is compressed and uses less bandwidth (8kbps per call) and is typically used across the WAN.
- Codec selection is controlled by the Region settings configured in CM Administration.



Most VoIP deployments assume limited bandwidth between sites over the WAN:

- Slow links (T1 = 1.54 Mbps)
- Existing data traffic may already take up most of the bandwidth
- QoS is required to ensure voice traffic gets priority
- Codec selection is important to minimize bandwidth use while maximizing voice quality

Within a single site (HQ, Regional Office, Branch), bandwidth is typically plentiful

- LAN speed (10/100/1000 Mbps)
- QoS may still be needed
- Codec selection is important to provide good voice quality and compatibility with other devices or applications (Messaging, Conferencing)

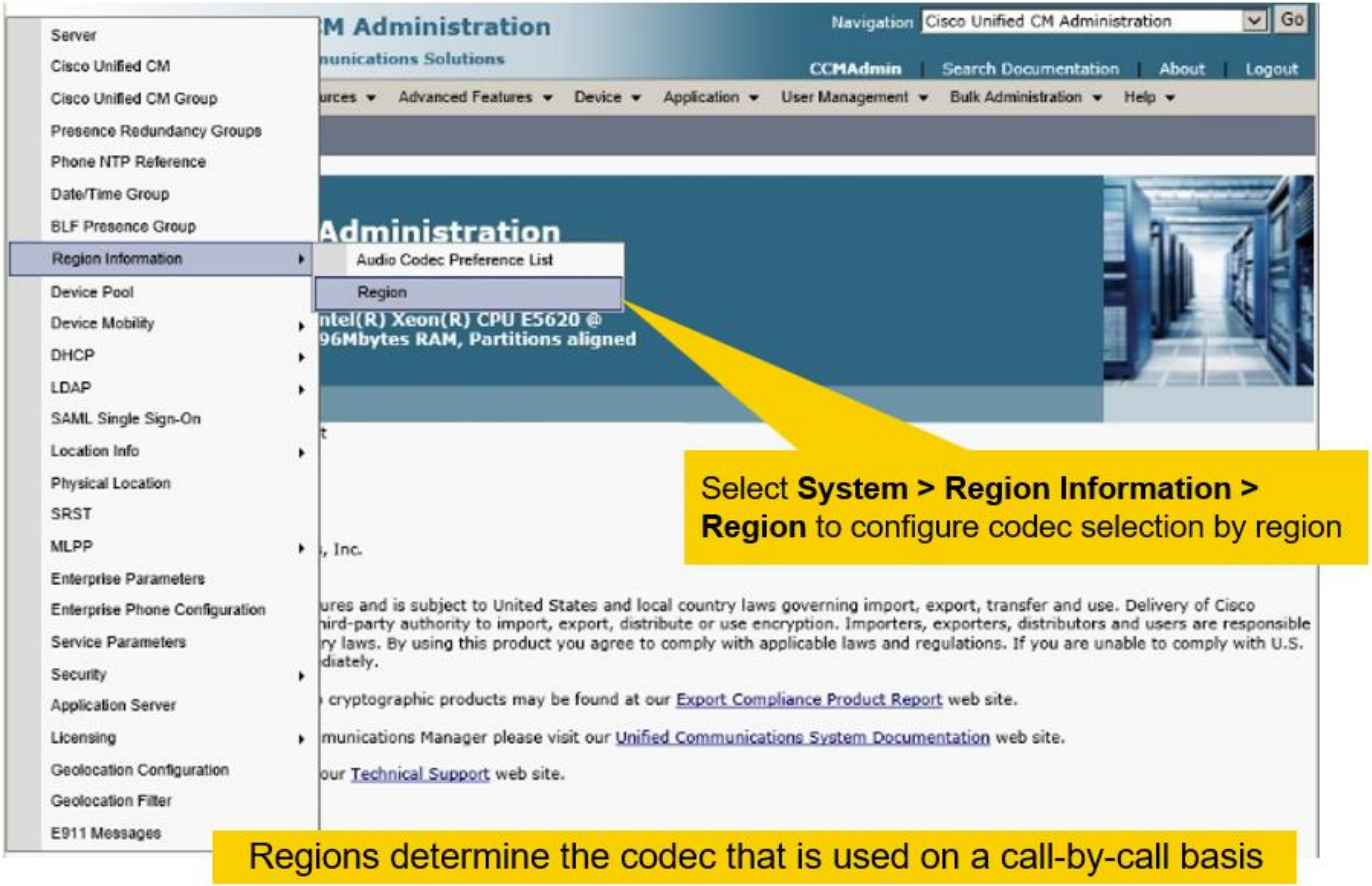
Bandwidth in CUCM

Codec bandwidth utilization is calculated by adding the codec payload bitrate to the header overhead:

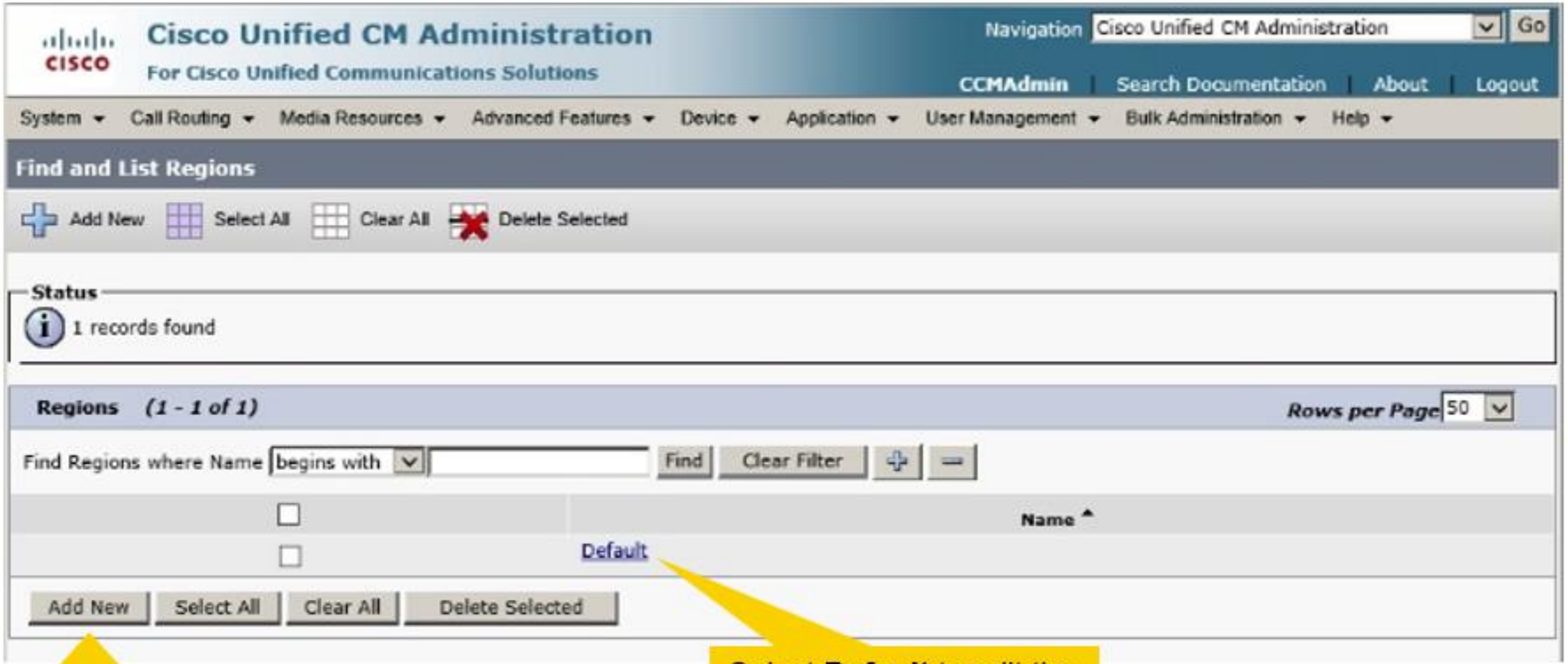
- Payload + Overhead = 16k bps

Codec	Payload	+Overhead
G.711	64 Kbps	80 Kbps
G.729	8 Kbps	24 Kbps

Regions and Codecs Configuration



Regions and Codecs Configuration (Cont.)

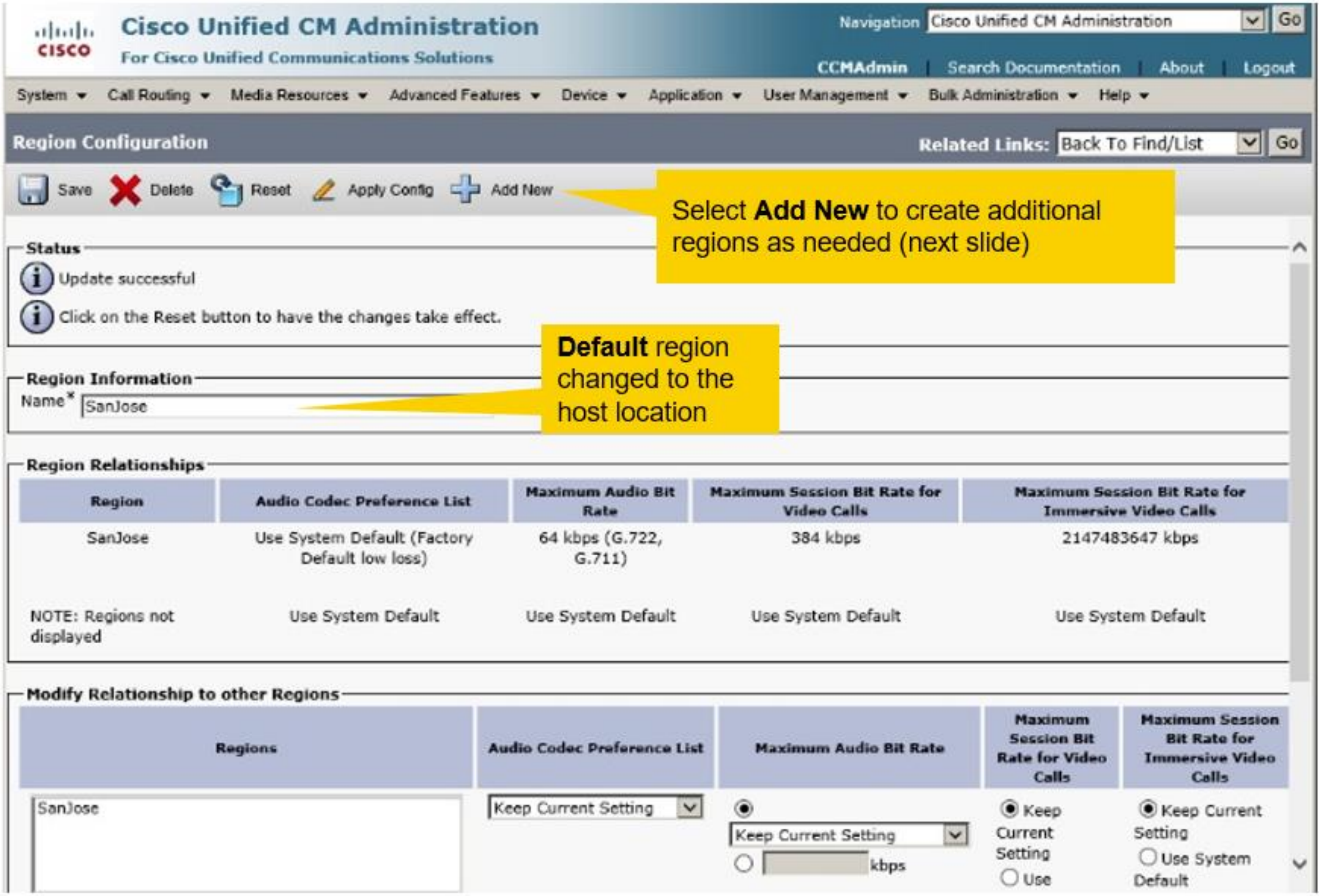


Click **Add New** to create a new region

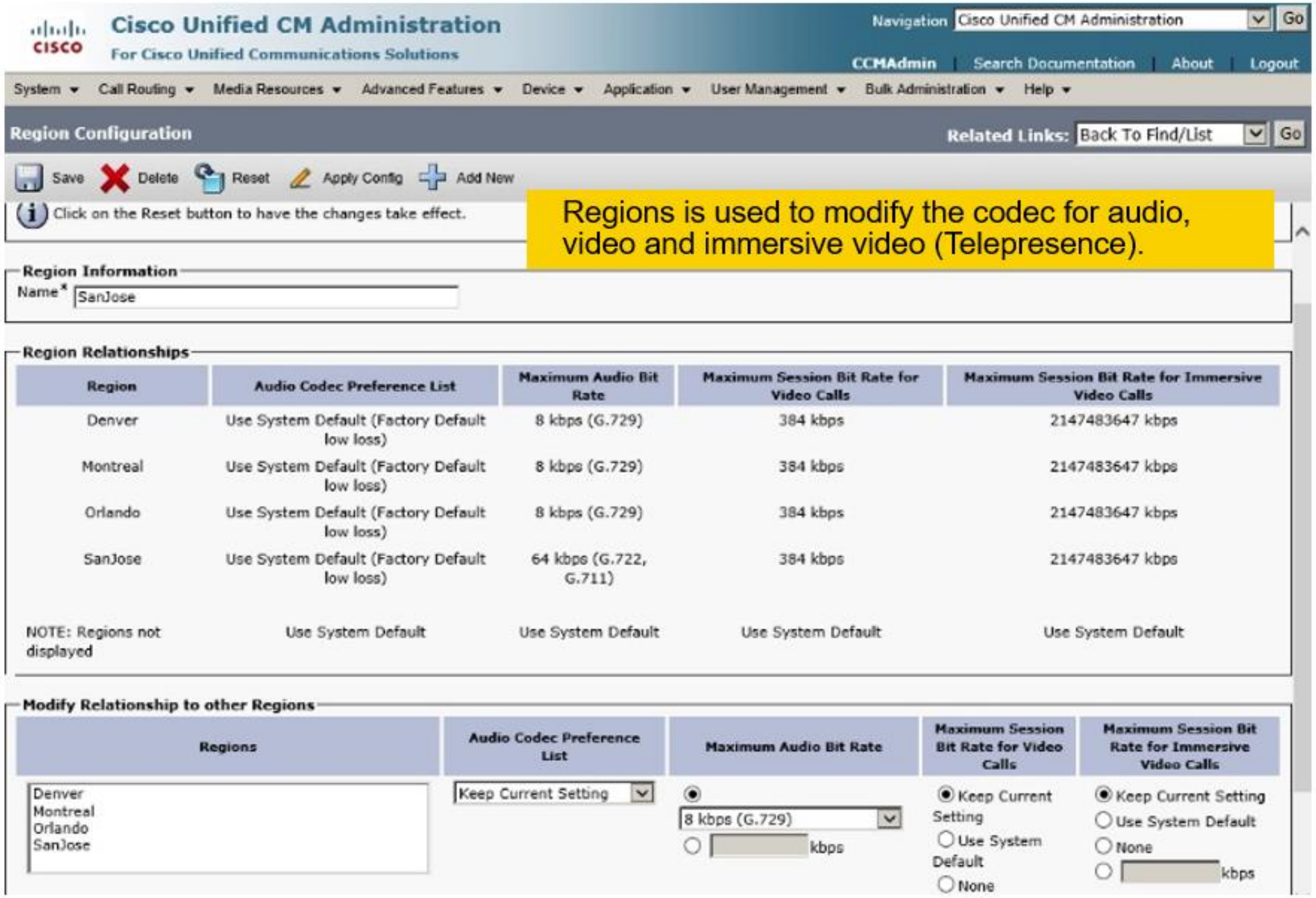
Select **Default** to edit the Default region (next slide)

Click **Default**, to view the Default Region configuration
The Default region is created automatically

Regions and Codecs Configuration (Cont.)



Regions and Codecs Configuration (Cont.)



Regions is used to modify the codec for audio, video and immersive video (Telepresence).

athena

Codec Selection in IOS Devices

Configuring Codec List

```
router(config)#
```

```
voice class codec class_tag
```

- Creates a codec voice class

```
router(config-class)#
```

```
codec preference value codec-type [mode frame-size][bytes payload-size]
```

- Configures the codec voice class with codecs and their preferences
- Mode and frame size apply to iLBC:
  - 20: 20-ms frames for 15.2 kb/s bit rate (default)
  - 30: 30-ms frames for 13.33 kb/s bit rate
- Payload size: voice payload of each frame
  - Values depend on the codec type
- Additional options exist for GSMAMR-NB codec

Codec-Related Dial Peer Configuration

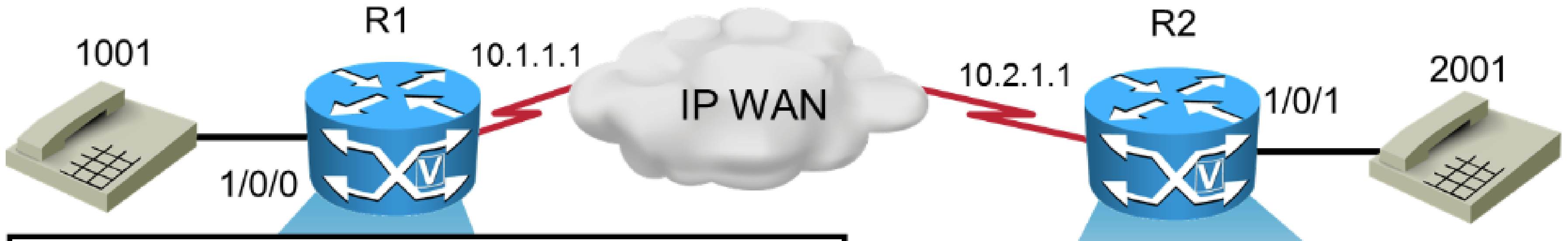
```
router(config-dial-peer)#
voice-class codec class_tag
```

- Assigns codec voice class to dial peer (multiple codec option)

```
router(config-dial-peer)#
codec {codec [bytes payload-size] | transparent} [fixed-bytes]
```

- Defines an individual codec on a dial peer
- **payload size**: voice payload of each frame
  - Values depend on the codec type
- **transparent**: enables codec capabilities to be passed transparently between endpoints in a Cisco Unified Border Element
- **fixed-bytes**: codec byte size is fixed and nonnegotiable
- Default: g729r8, 20-byte payload

Codec Configuration Example




```
voice class codec 100
 codec preference 1 g711alaw
 codec preference 2 g711ulaw bytes 80
 codec preference 3 g723ar53
 codec preference 4 g723ar63 bytes 144
 codec preference 5 g723r53
 codec preference 6 g723r63 bytes 120
 codec preference 7 g726r16
 codec preference 8 g726r24
 codec preference 9 g726r32 bytes 80
 codec preference 10 g728
 codec preference 11 g729br8
 codec preference 12 g729r8
dial-peer voice 1 voip
 destination-pattern 200.
 session target ipv4:10.2.1.1
 voice-class codec 100
```

```
dial-peer voice 4 voip
 destination-pattern 100.
 session target ipv4:10.1.1.1
```

- Result: g729r8, 20 bytes

Summary

- Voice Packetization
- VoIP Media Transmission
- VAD
- Codec Quality
- Overhead
- Codec Selection in CUCM
- Codec Selection in IOS Devices
- Closing & Wrap Up
  Your Questions and Feedback

Questions

Thank you for attending.

If you have any additional questions, or would like to learn more about our Athena program, please email...
pka@skyline-ats.com

Skyline
Advanced Technology Services


Skyline
Advanced Technology Services
www.skyline-ats.com