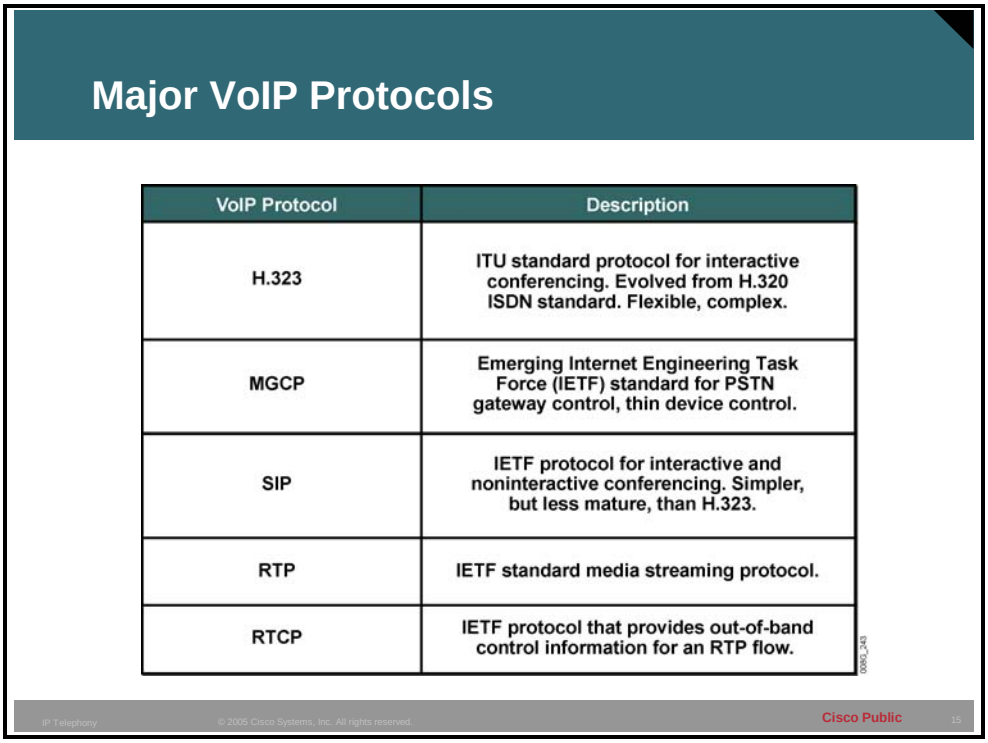
Encapsulating Voice in IP Packets

Major VoIP Protocols

This topic defines the major VoIP protocols and matches them with the seven layers of the OSI model.



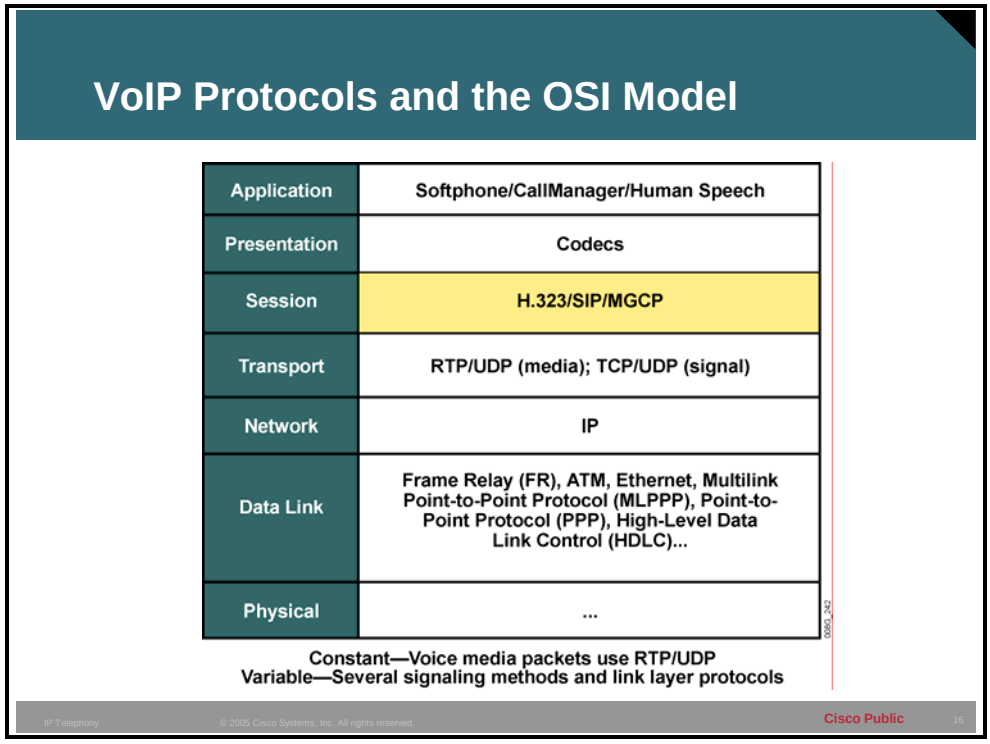
The major VoIP protocols include the following:

- **H.323:** An ITU standard protocol for interactive conferencing. The ITU standard protocol was originally designed for multimedia in a connectionless environment, such as a LAN. The H.323 is an umbrella of standards that defines all aspects of synchronized voice, video, and data transmission. H.323 defines end-to-end call signaling.
- **MGCP:** An emerging standard for PSTN gateway control or thin device control. Specified in RFC 2705, MGCP defines a protocol to control VoIP gateways connected to external call-control devices, referred to as call agents. MGCP provides the signaling capability for less expensive edge devices, such as gateways, that may not contain a full voice-signaling stack, such as H.323. In essence, any time an event such as off hook occurs at the voice port of a gateway, the voice port reports that event to the call agent. The call agent then signals that device to provide a service, such as dial-tone signaling.
- **SIP:** A detailed protocol that specifies the commands and responses to set up and tear down calls. It also details features such as security, proxy, and transport (TCP or UDP) services. SIP and its partner protocols, Session Announcement Protocol (SAP) and Session Description Protocol (SDP), provide announcements and information about multicast sessions to users on a network. SIP defines end-to-end call signaling between devices. SIP is a text-based protocol that borrows many elements of HTTP, using the same transaction request and response model, and similar header and response codes. It also adopts a

modified form of the URL-addressing scheme used within e-mail that is based on Simple Mail Transfer Protocol (SMTP).

- **RTP:** An Internet Engineering Task Force (IETF) standard media-streaming protocol. RTP carries the voice payload across the network. RTP provides sequence numbers and time stamps for the orderly processing of voice packets.
- **RTCP:** Provides out-of-band control information for an RTP flow. Every RTP flow has a corresponding RTCP flow that reports statistics on the call. RTCP is used for QoS reporting.

 Copyright © 2005, Cisco Systems, Inc.



Example: VoIP and the OSI Model

Successfully integrating connection-oriented voice traffic in a connectionless-oriented IP network requires enhancements to the signaling stack. In some ways, the user must make the connectionless network appear more connection-oriented.

Applications such as Cisco IP Softphone and Cisco CallManager provide the interface for users to originate voice at their PCs or laptops and convert and compress it before passing it to the network. If a gateway is used, a standard telephone becomes the interface to users, so human speech is the application.

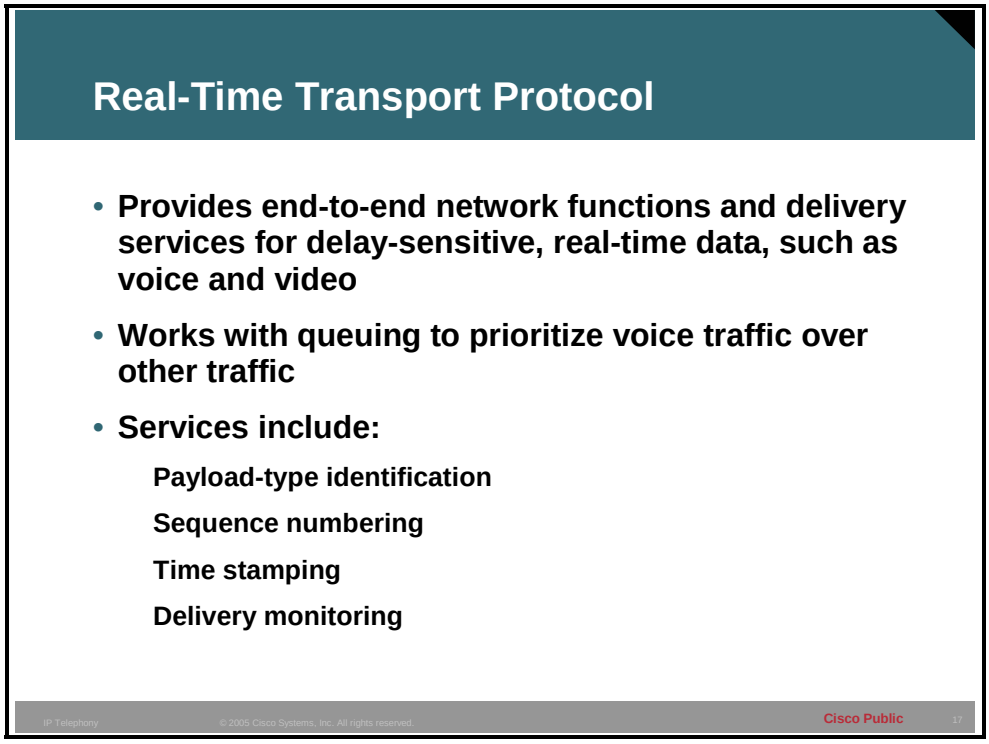
Codecs define how the voice is compressed. The user can configure which codec to use or a codec is negotiated according to what is available.

One of the constants in VoIP implementation is that voice uses RTP inside of UDP to carry the payload across the network. Because IP voice packets can reach the destination out of order and unsynchronized, the packets must be reordered and resynchronized before playing them out to the user. Since UDP does not provide services such as sequence numbers or time stamps, RTP provides sequencing functionality.

The variables in VoIP are the signaling methods used. H.323 and SIP define end-to-end call-signaling methods. MGCP defines a method to separate the signaling function from the voice call function. MGCP uses a call agent to control signaling on behalf of the endpoint devices, such as gateways. The central control device participates in the call setup only. Voice traffic still flows directly from endpoint to endpoint.

RTP and RTCP

This topic describes the functions of RTP and RTCP as they relate to the VoIP network.



RTP provides end-to-end network transport functions intended for applications transmitting real-time requirements, such as audio and video. Those functions include payload-type identification, sequence numbering, time stamping, and delivery monitoring.

RTP typically runs on top of UDP to use the multiplexing and checksum services of that protocol. Although RTP is often used for unicast sessions, it is primarily designed for multicast sessions. In addition to the roles of sender and receiver, RTP also defines the roles of translator and mixer to support the multicast requirements.

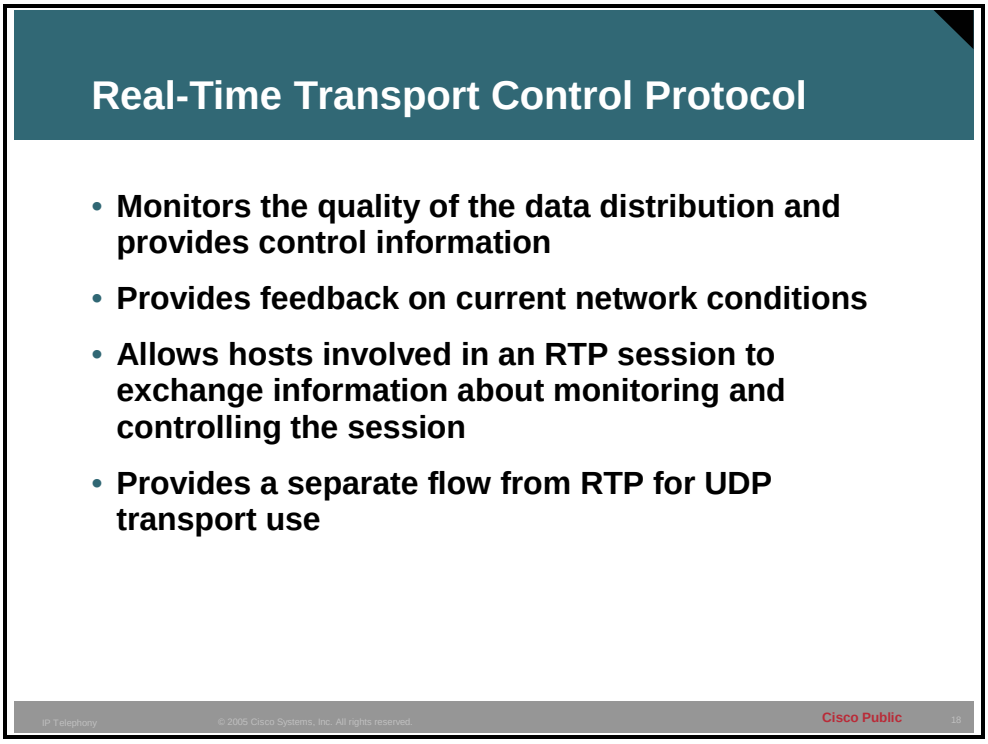
RTP is a critical component of VoIP because it enables the destination device to reorder and retime the voice packets before they are played out to the user. An RTP header contains a time stamp and sequence number, which allows the receiving device to buffer and remove jitter and latency by synchronizing the packets to play back a continuous stream of sound. RTP uses sequence numbers to order the packets only. RTP does not request retransmission if a packet is lost.

Example: RTP Application

As voice packets are placed on the network to reach a destination, they may take one or more paths to reach their destination. Each path may have a different length and transmission speed, resulting in the packets being out of order when they arrive at their destination. As the packets were placed on the wire at the source of the call, RTP tagged the packets with a time stamp and sequence number. At the destination, RTP can reorder the packets and send them to the digital signal processor (DSP) at the same pace as they were placed on the wire at the source.

Note For more information on RTP, refer to RFC 1889.

 Copyright © 2005, Cisco Systems, Inc.



RTCP monitors the quality of the data distribution and provides control information. RTCP provides the following feedback on current network conditions:

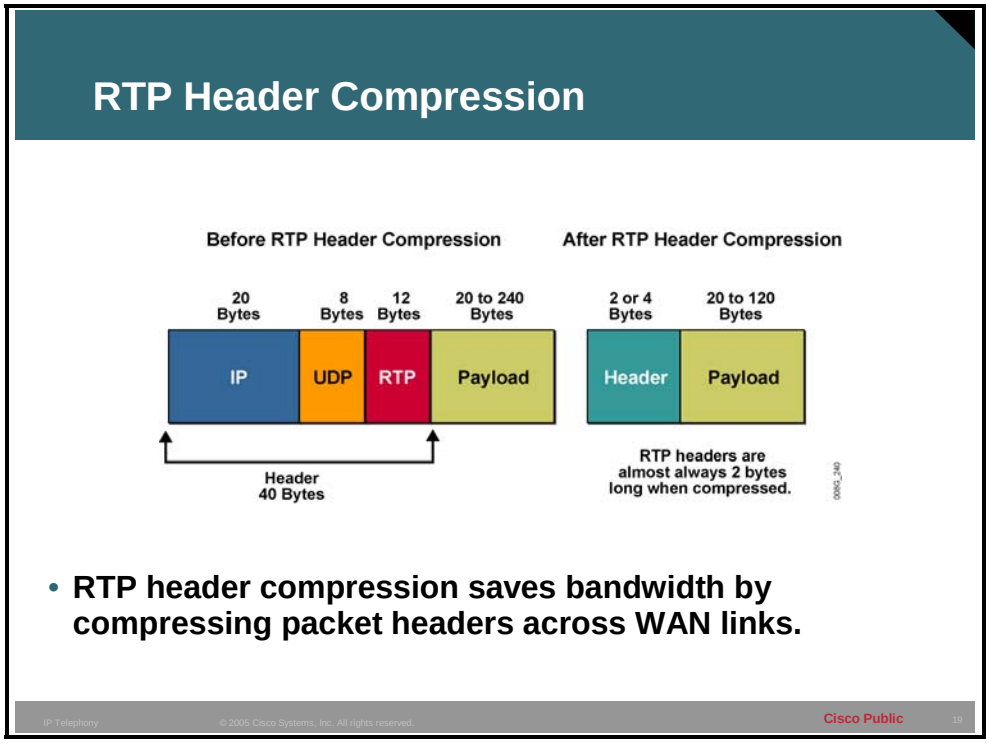
- RTCP provides a mechanism for hosts involved in an RTP session to exchange information about monitoring and controlling the session. RTCP monitors the quality of elements such as packet count, packet loss, delay, and interarrival jitter. RTCP transmits packets as a percentage of session bandwidth, but at a specific rate of at least every 5 seconds.
- The RTP standard states that the Network Time Protocol (NTP) time stamp is based on synchronized clocks. The corresponding RTP time stamp is randomly generated and based on data-packet sampling. Both NTP and RTP are included in RTCP packets by the sender of the data.
- RTCP provides a separate flow from RTP for transport use by UDP. When a voice stream is assigned UDP port numbers, RTP is typically assigned an even-numbered port and RTCP is assigned the next odd-numbered port. Each voice call has four ports assigned: RTP plus RTCP in the transmit direction and RTP plus RTCP in the receive direction.

Example: RTCP Application

Throughout the duration of each RTP call, the RTCP report packets are generated at least every 5 seconds. In the event of poor network conditions, a call may be disconnected due to high packet loss. When viewing packets using a packet analyzer, a network administrator could check information in the RTCP header that includes packet count, octet count, number of packets lost, and jitter. The RTCP header information would shed light on why the calls were disconnected.

Reducing Header Overhead with CRTP

This topic describes how IP voice headers are compressed using CRTP.



Given the number of protocols that are necessary to transport voice over an IP network, the packet header can be large. You can use CRTP headers on a link-by-link basis to save bandwidth.

Using CRTP compresses the IP/UDP/RTP header from 40 bytes to 2 bytes without UDP checksums and from 40 bytes to 4 bytes with UDP checksums. RTP header compression is especially beneficial when the RTP payload size is small; for example, with compressed audio payloads between 20 and 50 bytes.

In addition, CRTP works on the premise that most of the fields in the IP/UDP/RTP header do not change, or that the change is predictable. Static fields include source and destination IP address, source and destination UDP port numbers, as well as many other fields in all three headers. For those fields where the change is predictable, the CRTP process is illustrated in the following table:

CRTP

Stage	What Happens
The change is predictable.	The sending side tracks the predicted change.
The predicted change is tracked.	The sending side sends a hash of the header.
The receiving side predicts what the constant change is.	The receiving side substitutes the original stored header and calculates the changed fields.
There is an unexpected change.	The sending side sends the entire header without compression.

 Copyright © 2005, Cisco Systems, Inc.

CRTP Packet Components

In a packet voice environment when speech samples are framed every 20 ms, a payload of 20 bytes is generated. Without CRTP, the total packet size includes the following components:

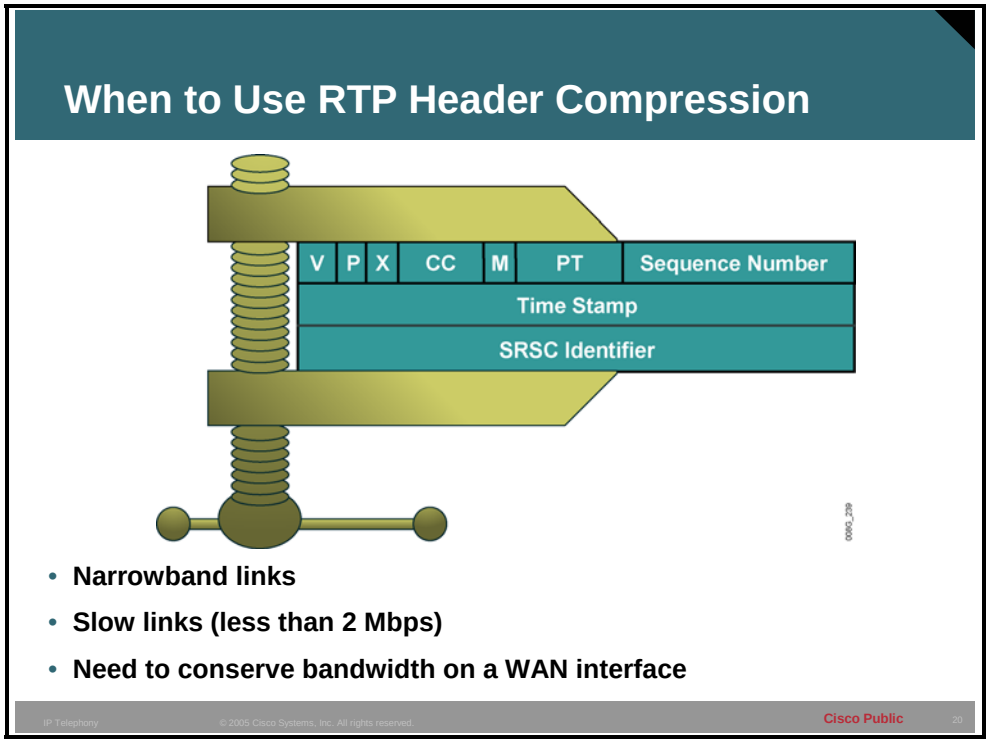
- IP header (20 bytes)
- UDP header (8 bytes)
- RTP header (12 bytes)
- Payload (20 bytes)

The header is twice the size of the payload: IP/UDP/RTP (20 + 8 + 12 = 40 bytes) versus payload (20 bytes). When generating packets every 20 ms on a slow link, the header consumes a large portion of bandwidth.

In the figure, RTP header compression reduces the header to 2 bytes. The compressed header is one tenth of the payload size.

When to Use RTP Header Compression

This topic describes when to use CRTP.



You must configure CRTP on a specific serial interface or subinterface if you have any of these conditions:

- Narrowband links
- Slow links (less than 2 Mbps)
- Need to conserve bandwidth on a WAN interface

Compression works on a link-by-link basis and must be enabled for each link that fits these requirements. You must enable compression on both sides of the link for proper results. Enabling compression on both ends of a low-bandwidth serial link can greatly reduce the network overhead if there is a significant volume of RTP traffic on that slow link.

Note	Compression adds to processing overhead. You must check resource availability on each device prior to turning on RTP header compression.

Example: Applying CRTP

If you want the router to compress RTP packets, use the **ip rtp header-compression** command. The **ip rtp header-compression** command defaults to active mode when it is configured. However, this command provides a passive mode setting in instances where you want the router to compress RTP packets *only* if it has received compressed RTP on that interface. When applying to a Frame Relay interface, use the **frame-relay ip rtp header-compression** command.

 Copyright © 2005, Cisco Systems, Inc.

By default, the software supports a total of 16 RTP header compression connections on an interface. Depending on the traffic on the interface, you can change the number of header compression connections with the **ip rtp compression-connections** *number* command.

Note	Do not use CRTP if you have high-speed interfaces or links faster than 2 Mbps.