

# Voice Over IP

**Matt Holbert, CCIE #6334**

# Table of Contents

**INTRODUCTION .....5**

**BACKGROUND .....6**

    CIRCUIT -SWITCHED VS. PACKET -SWITCHED..... 6

    ADVANTAGES & DISADVANTAGES OF A CONVERGED NETWORK..... 7

    OVERCOMING THE DISADVANTAGES ..... 8

**VOICE PORTS .....9**

    ANALOG..... 9

*Basic Configuration*..... 9

*Example*..... 10

    DIGITAL ..... 11

*Basic Configuration*..... 11

*Example*..... 12

**“VOICE OVER PACKET” TECHNOLOGY .....13**

    DIAL MAPPING..... 13

*Dial Peers*..... 13

*Call Legs*..... 13

*Dial Plans*..... 14

*POTS Dial Peer Configuration* ..... 14

*Direct-Inward Dialing*..... 15

*Network Dial Peer Configuration*..... 16

    SESSION INITIATION..... 16

*H.323*..... 17

    VOICE-ENCODING ..... 17

*Pulse-Code Modulation*..... 17

*Adaptive-Differential Pulse Code Modulation (ADPCM)*..... 18

*Code-Excited Linear Prediction (CELP)*..... 19

*Configuration*..... 19

    ENCAPSULATION ..... 20

    TRANSPORT ..... 20

*Voice over HDLC* ..... 20

*VoHDLC Example* ..... 20

*Voice over IP* ..... 21

*VoIP Example*..... 22

*Voice over Frame Relay* ..... 23

*VoFR Example*..... 23

*Voice over IP over Frame Relay*..... 24

*VoIPoFR Example* ..... 24

*Voice over ATM*..... 25

*VoATM Example* ..... 25

**QUALITY OF SERVICE.....27**

    DELAY BUDGET ..... 27

    HEADER COMPRESSION..... 28

*Compressed RTP (cRTP)*..... 28

    QUEUEING ..... 30

*FIFO Queuing*..... 30

*Weighted Fair Queuing*..... 30

*Priority Queuing*..... 31

*Custom Queuing*..... 32

PACKET CLASSIFICATION.....	33
<i>IP Precedence</i> .....	33
<i>RSVP</i> .....	34
<i>IP RTP Reserve</i> .....	35
<i>IP RTP Priority</i> .....	36
TRAFFIC POLICING.....	36
<i>CAR</i> .....	36
TRAFFIC SHAPING.....	38
<i>GTS</i> .....	38
<i>FRTS</i> .....	39
<b>SUMMARY.....</b>	<b>43</b>

## Table of Figures

Figure 1: TDM Switching of T1s .....	6
Figure 2: Toll Bypass .....	7
Figure 3: Voice over Packet Process .....	13
Figure 4: Call Legs .....	14
Figure 5: Sampling and Quantizing .....	18
Figure 6: Simple VoHDL Example .....	21
Figure 7: Simple VoIP Example .....	22
Figure 8: Simple VoFR Example .....	23
Figure 9: Simple VoIPoFR Example .....	24
Figure 10: Simple VoATM Example .....	25
Figure 11: Compressed RTP Comparison .....	29
Figure 12: FIFO Queuing .....	30
Figure 13: Weighted Fair Queuing .....	31
Figure 14: Priority Queuing .....	31
Figure 15: Custom Queuing .....	32

---

## INTRODUCTION

---

The convergence of telephony services into data networks has the ability to dramatically change the business communication of the world by delivering information more efficiently than today's dual-network approach. A converged voice, video, and data network improves customer care capabilities, increases employee productivity, and enhances corporate agility.

The primary goal of this paper is to help prepare Cisco Certified Internetwork Expert (CCIE) candidates for the voice section of the practical, hands-on laboratory exam. However, any curious network engineer or systems administrator with even a limited amount of networking experience should be able to use this paper as a guide. The paper includes many examples for possible use as design templates.

The paper begins with a (very) brief history of voice technology. The paper then discusses the two types of voice ports, analog and digital. Next, discussion turns to the technology and configuration processes of voice over packet technology. Finally, the paper presents several protocols, with examples, for enabling Quality of Service (QoS) for the voice network. Several configuration examples accompany the text. All examples use version 12.0 of the Cisco Internetwork Operating System (IOS).

All IOS command explanations use the Cisco conventions as shown in Table 1-Command Conventions.

Command Convention	Explanation
normal	Indicates commands and keywords that are entered literally as shown.
<i>italics</i>	Indicates arguments for which you supply values.
[command]	Keywords or arguments that appear within square brackets are optional.

**Table 1: Command Conventions**

---

## BACKGROUND

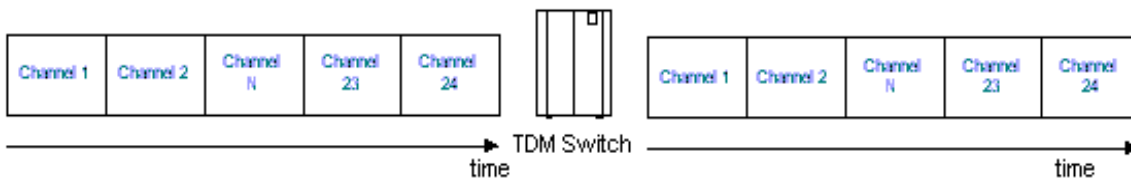
---

Alexander Graham Bell invented the telephone on March 10, 1876 in Boston, Massachusetts. As the invention caught on, direct wire connections provided connectivity between callers. This soon became impractical as the number of telephones began to increase, thus the profession of telephone operator emerged. Now a single wire connected each house to a switchboard, where the operator manually connected calls. Later, the invention of the circuit-switch would replace the operator.

Today's voice networks are digital instead of analog. One such digital circuit, known as a T1, carries up to 24 simultaneous Time-Division Multiplexed (TDM) calls, or channels. To place a call, the voice network will reserve a channel through the network for the call. This is termed a *circuit-switched* network because the network switches individual channels for the duration of the call.

Figure 1 is a simple explanation of circuit-switching. In the scenario, two T1s terminate into a TDM switch. A call on channel 1 of the left T1 may switch to channel 23 of the right T1. Realistically, the TDM switch terminates several T1s to connect geographically dispersed areas. For example, the TDM switch may link Chicago, Los Angeles, and New York.

The TDM switches use a signaling system for call setup to establish a channel through the network for the call. This same signaling system releases the dedicated channel upon hang-up.



**Figure 1: TDM Switching of T1s**

Data networks are packet-switched. *Packet-switched* networks require placing a tag known as a header onto the data, allowing the data to be routed through the network to the destination. Each switch throughout the network examines the header, chooses the next hop, and then transmits the data to the next hop. This continues hop-by-hop until the data is delivered to the ultimate destination.

---

## CIRCUIT-SWITCHED VS. PACKET-SWITCHED

---

The transit delay of packets through a circuit-switched network is deterministic. After the call setup, the delay is always the same because the system reserves the channels for the entire length of the call. Packet-switched networks encounter a variable delay since individual packets may take different paths or reside for a longer time on a router before being routed.

Circuit-switched networks require full use of the available channel bandwidth throughout the call. Packet-switched networks, on the other hand, only consume bandwidth as needed. Thus, the bandwidth efficiency of a packet-switched network is efficient compared to the inefficiency of a circuit-switched network. Table 2 summarizes the differences between the two switching methods.

	Circuit-Switched	Packet-Switched
<b>Transit Delay</b>	Deterministic	Variable
<b>Bandwidth Consumption</b>	Full	Variable
<b>Bandwidth Efficiency</b>	Inefficient	Efficient

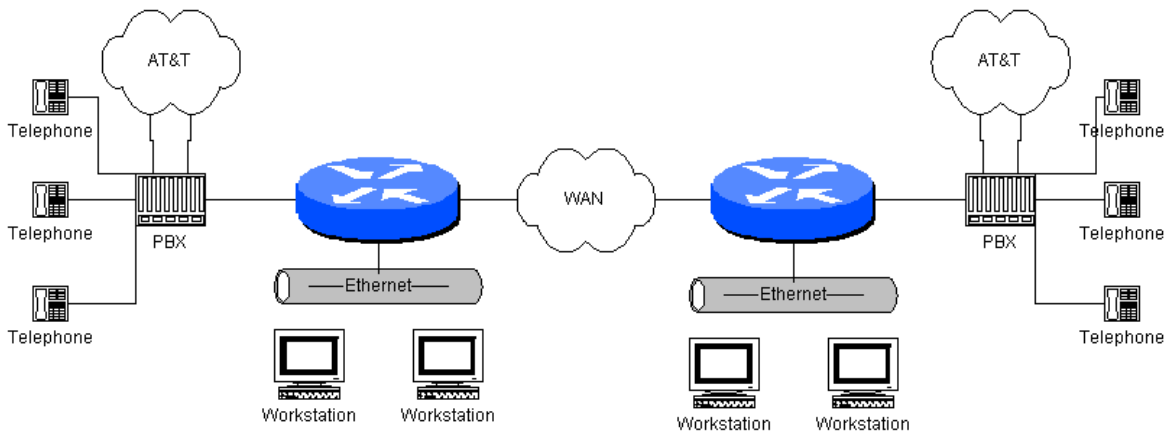
**Table 2: Switching Comparison**

Traditionally, voice and data networks were segmented into two distinct networks. After all, voice is time-sensitive and requires constant bandwidth. Data, on the other hand, is not relatively time-sensitive and typically consumes a variable amount of bandwidth. It is for these reasons that voice has been typically circuit-switched and data has been packet-switched.

Recently, however, advances in Application-Specific Integrated Circuits (ASICs) installed in Digital Signal Processors (DSPs) have allowed a transition to a converged network of voice and data over a packet-switched network.

## ADVANTAGES & DISADVANTAGES OF A CONVERGED NETWORK

The advantages of integrating the voice network with the data network are enormous. Most importantly is the cost savings of *toll bypass*, or routing calls entirely over data networks. Figure 2 depicts two sites connected by a data network. If calls between the two sites traverse the data network (WAN), the calls avert long-distance charges. This is toll-bypass.



**Figure 2: Toll Bypass**

Almost as important is the elimination of maintaining two separate networks. Only one physical network is required, thus daily management and maintenance of the network requires only one team. And for the business world, enhanced services such as voice-enabled Web sites, document sharing, and unified messaging are available with a voice and data network that were previously unavailable.

The disadvantages of a converged network are stalling the deployment of these networks. Despite the savings, few will sacrifice cost if the same level of service of traditional telephony cannot be guaranteed. Another disadvantage is inherent on the dependency of

voice on the data network. That is, if the data network is non-operational, so may be the voice network.

### **Advantages**

- **Toll bypass**
- **One network to manage and maintain**
- **Enhanced services**

### **Disadvantages**

- **Quality of Service not yet optimized**
- **Dependency on data network**

---

## **OVERCOMING THE DISADVANTAGES**

Many leaps have been accomplished regarding these disadvantages, thus voice and data integrated networks have begun to spread in popularity. The Quality of Service (QoS) over private networks can be stipulated in the Service Level Agreement (SLA), but with this QoS comes price. Regardless, a private network is the most popular transport of a converged network. The public Internet also has begun to adapt to this request for differentiated service for voice and data. Some Internet Service Providers (ISPs) now offer SLAs for traffic routed entirely within their network making the transport of voice over the Internet a possibility. Until the entire Internet can agree on standards to implement this badly needed QoS, most will continue to use private networks.

Because the voice network is now dependent upon the data network, engineers must design converged networks with redundancy in mind. This is costly, as designers must purchase multiple pieces of equipment and multiple links. It is required, however, to ensure that non-operational equipment does not bring the data and/or voice network as a whole down.

---

## VOICE PORTS

---

Cisco "Voice over Packet" technology supports both analog and digital voice ports. The voice ports connect to the router via Voice Interface Cards (VICs) installed within Voice Network Modules (VNMs). Some voice port types allow connecting standard analog phones, while other types connect to a PBX or the PSTN central office. Care must be taken to ensure the voice port is used correctly. Plugging equipment into the wrong type of voice port can damage equipment.

---

### ANALOG

Analog voice ports can connect directly to basic telephone equipment, such as analog telephones or faxes, to a PBX, to the PSTN central office, or to PBX trunk lines.

Voice signaling communicates attributes—such as on-hook, off-hook, call setup, dialed digit, etc.—between telephony components. The type of port used specifies the type of signaling used. Analog voice ports support three basic voice-signaling types.

- **Foreign Exchange Station (FXS).** This is the type of interface commonly seen in the CCIE lab. This RJ-11 interface allows connections to basic telephone equipment, such as analog telephones or faxes, or PBXs. It supplies ring, voltage, and dial tone to the telephone equipment. It can also serve as a single analog line between the PBX and the router.
- **Foreign Exchange Office (FXO).** The FXO interface is a RJ-11 connector that allows a connection to the PSTN central office or to a standard PBX interface. The connection to the central office allows both inbound and outbound calls to/from the PSTN. Use this type of interface in scenarios where the distance between the PBX and the router is greater than that allowable by an FXS interface.
- **E&M.**<sup>1</sup> This RJ-48 connector interface allows connections to PBX trunk lines (tie lines). It is a signaling technique for 2-wire and 4-wire telephone and trunk interfaces.

The type of voice port required depends primarily on the device connection. For example, if the PBX has a FXO port, you need a FXO port for the router.

### BASIC CONFIGURATION

Voice-port commands define the voice-port signaling type, as well as any characteristics required such as dial-type, ring frequency, etc. The default voice-port configuration is typically sufficient to configure FXO and FXS ports to transport voice over the IP network. E&M ports, on the other hand, usually require particular values configured dependent upon the specifications of the PBX devices in the telephony network.

---

<sup>1</sup> Several theories exist explaining where the acronym E&M originates. The two most popular are the "Ear and Mouth" interface and "recEive and transMit" interface. Another theory evolves from Earth and Magnet, the earth representing ground and the magnet representing the signal.

Voice-port configuration requires entering voice-port configuration mode for the chosen voice-port. Use the typical IOS interface numbering to number the voice-ports. From configuration mode, the following command enters voice-port configuration mode.

```
Router(config)#voice-port slot-number/subunit-number/port
```

Rather than specify in detail the particulars of every possible configuration command, it is best to simply outline some of the major commands. Enter each command presented in Table 3 in voice-port configuration mode.

Command	Default	Comment
dial-type {dtmf   pulse}	dial-type dtmf	(FXO ports only) Selects dial-type, either touch-tone (DTMF) or pulse.
signal {loop-start   ground-start}	signal loop-start	With the loop-start keyword, only one side of a connection can hang up. With ground-start signaling, both sides of a connection can place calls and hang up.
ring frequency {25   50}	ring frequency 25	(For FXS ports only) Selects ring frequency (in Hertz).
ring number <i>number</i>	ring number 1	(For FXO ports only) Maximum number of rings detected before answering a call.
connection {plar   trunk} <i>string</i>	No connection mode specified.	Sets up a connection mode for the voice port.

**Table 3: Analog Voice Port Configuration**

The last command presented in Table 3 is a very important one. A trunk connection is used when the port is to act as a medium between two PBXs. For example, assume you have two PBXs, one connected to a router's voice-port and the other connected to voice-port on another router. If you configure the two ports as trunks, it will appear to the PBXs that they have a direct connection.

PLAR, or Private-Line Auto Ringdown, automatically dials a preset number as soon as the phone is taken off-hook. This is synonymous with the bat cave phone. As soon as batman picks up the phone, the phone automatically places a call to Alfred the Butler.

Again, FXO and FXS ports typically require no configuration. The above commands are simply the most useful commands. Other fine-tuning commands are available, such as setting music thresholds, comfort noise levels, and descriptions.

### EXAMPLE

The example below is the configuration of two ports. The first one requires no special configuration. The second port includes a PLAR connection to another phone.

```
voice-port 1/0/0
```

```
!
voice-port 1/0/1
    connection plar 101
```

## DIGITAL

Digital voice ports can connect to a PBX or to the PSTN central office switch. In North America, the most popular digital interface is the T1. Some digital voice ports on Cisco routers have a built-in Channel Service Unit/Data Service Unit (CSU/DSU). When configuring a digital voice port, in reality the CSU/DSU is being configured.

### BASIC CONFIGURATION

Configuring a digital interface is straightforward. The configuration simply requires acquiring the proper line encoding, framing, signaling mode, clock source, and channel allocation for the digital line. Practically, many times the toughest part is acquiring the needed information from the PSTN or the PBX technician.

Digital voice-port configuration requires entering controller configuration mode for the chosen digital voice-port. From configuration mode, the following command enters controller configuration mode.

```
Router(config)#controller t1 slot-number/subunit-number/port
```

Enter controller configuration mode for each command presented in Table 4.

Command	Example	Comment
mode {atm, cas}	mode cas	Configure channel-associated signaling or ATM signaling. ATM signaling can only be used for WAN ports, not voice ports.
framing {sf, esf, crc4}	framing esf	Configures physical-layer framing.
linecode {b8zs, ami}	linecode b8zs	Configures physical-layer line coding.
clock source {line, internal}	clock source line	Sets the clock source for the interface.
voice-group <i>number</i> timeslots <i>slots</i> [ <i>type type</i> ]	voice-group 1 timeslots 1-8 type e&m	Configures a list of timeslots to form a CAS group for the T1/E1 line.

**Table 4: Digital Voice-Port Configuration**

Channelized T1s do require Channel Associated Signaling (CAS)—also known as robbed-bit signaling—to communicate signaling information to the PBX. The signaling protocol robs every sixth voice frame of a bit. This least-significant bit does not affect voice quality, but does allow insertion of the signaling information. When configuring MultiFlex Trunks

(MFTs)<sup>2</sup>, configuring the interface controller for CAS is not required on ISDN Primary Rate Interfaces (PRIs). This is because ISDN uses the out-of-band D channel for signaling.

### EXAMPLE

A typical configuration of a T1 connected to a PBX is below. The signaling uses robbed-bit signaling (CAS). The framing is extended super-frame (ESF), the linecode is bipolar eight with zero substitution (b8zs), and the PBX is supplying the clock. The configuration allows up to eight concurrent calls on channels 1 through 8. All other channels are unused.

```
controller t1 0
  mode cas
  framing esf
  linecode b8zs
  clock source line
  voice-group 1 timeslots 1-8 type e&m-immediate
```

---

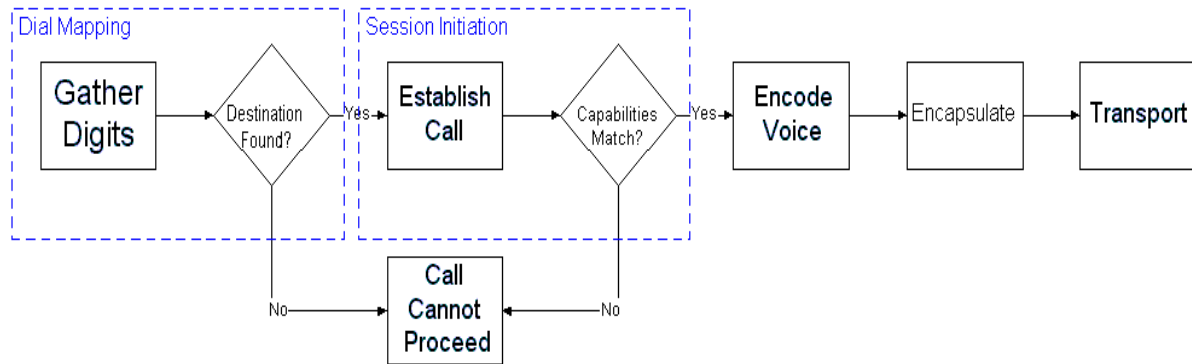
<sup>2</sup> MFTs allow voice and data over the same circuit. They are primarily used for connections to the service provider (WAN).

---

## “VOICE OVER PACKET” TECHNOLOGY

---

Outlined in Figure 3 is the complete process of voice over packet technology. The following sections explain in detail each individual procedure.



**Figure 3: Voice over Packet Process**

---

### DIAL MAPPING

When a user first picks up the phone handset, it signals the PBX or router of the off-hook condition. The PBX or the router's analog voice-ports supply a dial tone to the handset. The user then dials the phone number, and the router stores this information. This phone number maps to a destination using a *dial peer* that selects the corresponding *call leg*.

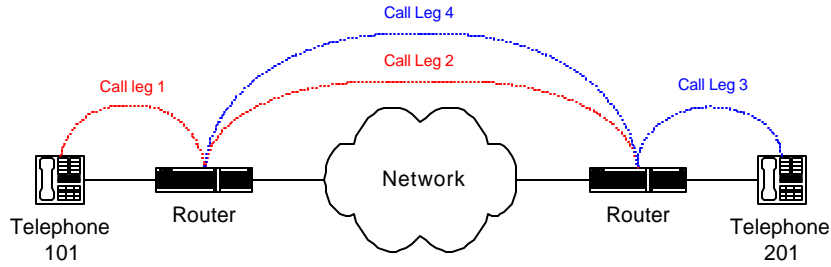
### DIAL PEERS

A dial peer defines the mapping between a phone number and the destination address. The mapping includes not only the destination address, but also certain attributes of the mapping, including Voice Activity Detection (VAD) possibilities, acceptable voice compression methods, and other attributes.

There are two types of dial peers, Plain Old Telephone System (POTS) and network. POTS dial peers represent a telephony device directly attached to the router. POTS dial peers include a phone number(s); a destination address is not needed because the peer is directly connected. Network dial peers represent a connection to another router. The network dial peer includes a destination phone number(s) and a destination address.

### CALL LEGS

A call leg is a connection between two telephony devices. There are four call legs for each established call, two for the originating router and two for the destination router. The first call leg is between the originating device—a handset, PBX, or the PSTN—and the router. The second call leg is between the source router and the destination router. The third and fourth call legs are the same as the first and second, only they are from the perspective of the destination router. Figure 4 below shows all four call legs.



**Figure 4: Call Legs**

Assume the left handset dials 201. The directly connected router receives the dialed digits, and looks in its dial peer mapping for the destination router. The digits are transported to the destination router, where another lookup occurs. This router uses a dial peer mapping to find the ultimate destination. The destination router also establishes its corresponding call legs. A bi-directional route now exists for the call between telephone 101 and telephone 201.

### DIAL PLANS

Dial peers and call legs have been defined. The actual configuration of the dial peers, however, has not been revealed.

Any efficient voice network design begins with a dial plan. A basic dial plan simply maps a telephone number, known as a destination pattern, to either a voice port or an address, known as a session target. A more detailed and practical dial plan also includes a peer number, type of peer, coding/decoding method, and QoS method, at the least. Table 5 is a sample dial plan.

Dial Peer	Type of Peer	Voice Port	Session Target	Destination Pattern	Extension	CODEC	QoS Method
1	POTS	2/0/0		+1301695....	5....		
2	POTS	2/0/1		+1301696....	6....		
3	VoIP		IPV4:207.197.132.201	+1301293....		G.729	cRTP
4	VoIP		IPV4:206.239.237.16	+1410715....		G.723a	RSVP
5	VoFR		DLCI:48	+1304262....		G.711	FRF.12

**Table 5: Sample Dial Plan**

Note that dial plans are router-centric. That is, each router must have its own dial plan consistent with the network dial plan as a whole.

### POTS DIAL PEER CONFIGURATION

POTS dial peer configuration requires entering dial peer configuration mode. The number of the dial peer simply must be unique. However, for management reasons I recommend to

sequentially number the peers. From configuration mode, the following command enters POTS dial peer configuration mode.

```
Router(config)#dial-peer voice number pots
```

Enter each command presented in Table 6 in POTS dial peer configuration mode.

Command	Example	Comment
destination-pattern [+]string[t]	destination-pattern +1301695...	Enters telephone number of the POTS peer.
port slot/sub-unit/port	port 2/0/0	Binds port to dial peer.
direct-inward dial	direct-inward dial	(Optional) Activates direct-inward dialing as specified in the Direct-Inward Dialing section below.
vad	vad	Enables voice-activity detection.

**Table 6: POTS Voice-Port Configuration**

The destination-pattern string is the phone number(s) of the POTS peer. Acceptable values are the numbers 0 through 9, the characters A through D, the \*, and the #. Also acceptable is a comma for a 1-second pause or a period for a single-digit wildcard. The optional leading + indicates a standard E.164 number. The optional trailing t indicates the router should wait for the inter-digit timeout before choosing a match. Only use the trailing t if variable-length dial strings are being used; otherwise, the inter-digit timeout is unnecessary. Table 7 gives several examples of destination-patterns.

Command	Description
destination-pattern +1234	Only the E.164 telephone number 1234
destination-pattern 123.	The number 123, followed by any number. As soon as the fourth digit is dialed, the match occurs and the call is routed. Other digits after the fourth are not used.
destination-pattern 1...t	Any four-digit number beginning with 1. If a fifth digit is dialed before the inter-digit timeout expires, this destination-pattern does not match.

**Table 7: Destination-Pattern Examples**

## DIRECT-INWARD DIALING

Without Direct-Inward Dialing (DID), when a call arrives on the router, the router presents a dial tone to the caller and collects digits until it can identify the destination dial peer. After

dial peer identification, the router forwards the call to the next call leg toward the destination.

Cases exist where it might be necessary for the router to use the called-number to find a dial peer for the outgoing call leg. With DID, the router does not present a dial tone to the caller and does not collect digits. It forwards the call directly to the configured destination. DID enables the router to match the called-number with a dial peer and then directly place the outbound call.

## NETWORK DIAL PEER CONFIGURATION

Dial peer configuration requires entering dial peer configuration mode for the type of dial peer used. The types available are VoHDLC, VoIP, VoFR, and VoATM. The peer number must be unique to both voice port peers and voice over peers. From configuration mode, the following command enters dial peer configuration mode.

```
Router(config)#dial-peer voice number {vohdlc, voip, vofr, voatm}
```

Enter each command presented in Table 8 in dial-peer configuration mode.

Command	Example	Comment
destination-pattern <i>string</i>	destination-pattern +1301695....	Enters telephone number of the POTS peer.
session target ipv4: <i>address</i> - or - session target dns: <i>hostname</i>	session target ipv4:207.197.132.233 - or - session target dns:www.icscorp.com	Binds IP destination to dial peer.
session target <i>interface number</i>	session target serial0	Binds HDLC destination to dial peer.
session target <i>interface number dci</i>	session target serial0 33	Binds Frame Relay destination to dial peer.
session target atm <i>number pvc</i>	session target atm0 1	Binds ATM destination to dial peer.

**Table 8: VoIP Dial Peer Configuration**

For specific examples of Voice over HDLC, Voice over IP, Voice over Frame Relay, and Voice over ATM, see the Transport section of this document.

---

## SESSION INITIATION

So far, only a route exists for the call. Nothing has been mentioned of call establishment and teardown, exchange of capabilities, or call control. Several session layers protocols exist

for achieving this. The H.323 session layer protocol is the most popular simply because it has been around the longest. Others, such as the Session Initiation Protocol (SIP) and Media Gateway Control Protocol (MGCP), are becoming more popular because they are simpler and more robust.

Discussion revolves only around the H.323 protocol. The other protocols serve the same primary purpose.

### H.323

Every voice-enabled router is an H.323 agent. The agent first uses the H.225 and Q.931 protocols to establish a TCP call using TCP port 1720 to the destination agent. The agent then establishes another TCP reliable connection for an H.245 capabilities exchange. This negotiation primarily selects audio and/or video codecs. The Real-Time Protocol (RTP) uses UDP to transport the audio stream. RTP provides some connection-oriented attributes (such as sequencing) without unneeded overhead. The Real-Time Control Protocol (RTCP) is used for the control channel. RTCP provides the monitoring of real-time audio delivery that UDP cannot provide.

---

## VOICE-ENCODING

The process of converting voice, which is inherently analog, into a digital format for transportation across a digital network is known as voice encoding, or *digitization*. At the destination, this process reverses to convert the digital values back into analog voice. Several different methods of accomplishing this will be discussed.

### PULSE-CODE MODULATION

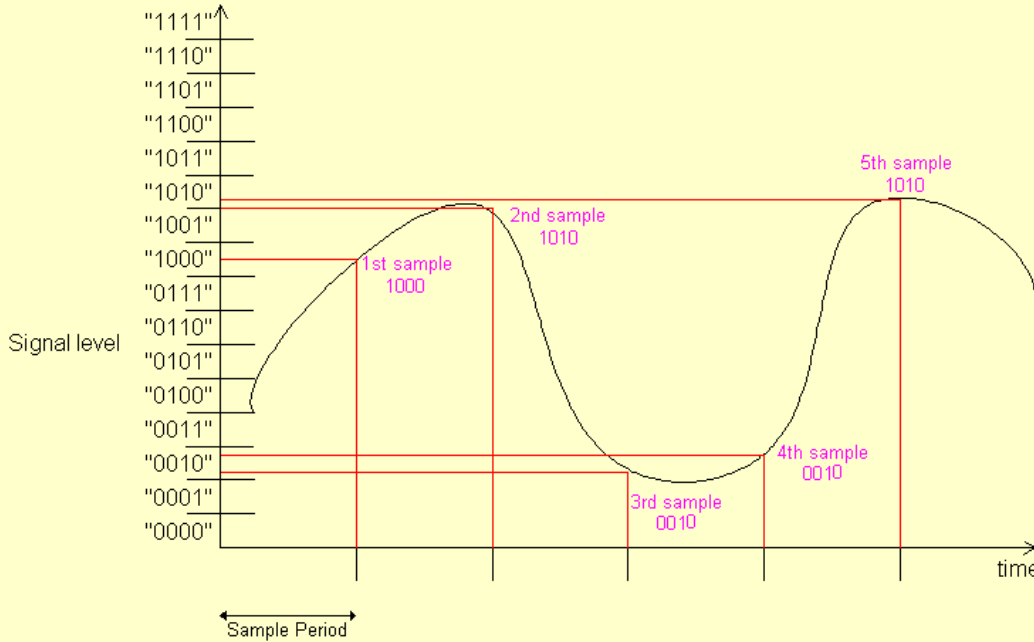
The simplest form of digitization is Pulse-Code Modulation (PCM). PCM is a two-step process. The first process is sampling, and the second process is quantizing.

*Sampling* works by measuring the amplitude of the analog signal at regular intervals. The more often the signal is sampled, the more accurate the signal is represented. The Nyquist Theorem states that to successfully reconstruct a waveform, sample the analog signal at twice the highest frequency. Most voice is below 4 kilohertz (kHz), so a sampling rate of 8000 times per second is the standard.

*Quantizing* works by assigning an integer value to this sampled analog signal. The more bits used for each sample, the more accurate the signal representation. The standard is 8 bits per sample. Taking 8000 samples per second and using 8 bits per sample yields the 64 kilobits per second throughput typical for voice.

$$8000 \frac{\text{samples}}{\text{second}} * 8 \frac{\text{bits}}{\text{sample}} = 64,000 \frac{\text{bits}}{\text{second}}$$

Figure 5 shows an example analog waveform with five samples. For each sample, the corresponding four-bit digital value is assigned. It is this digital sample that is transmitted.



**Figure 5: Sampling and Quantizing**

The International Telecommunications Union—Telecommunications Standardization Sector (ITU-T) has developed a standard based on 64 kbps PCM known as recommendation G.711.

**ADAPTIVE-DIFFERENTIAL PULSE CODE MODULATION (ADPCM)**

The popularity of cellular voice and wireless digital systems followed by voice and data integration created a need to decrease the required bandwidth for voice. One method for decreasing this bandwidth is Adaptive-Differential Pulse Code Modulation (ADPCM).

ADPCM allows the removal of redundant information by transmitting the difference between the current signal and the previous signal. For example, assume the previous signal was 45 decimal and the current signal is 58 decimal. Instead of transmitting 58 decimal, the difference of 13 decimal is transmitted. If the previous signal was 45 and the current signal is 32, the difference of -13 is transmitted.

To allow negative numbers representation in binary, a special convention need be adopted. The most common convention represents the most significant bit as a negative number. For example, the most significant bit of a four-bit number would have a value of -8 as opposed to 8. Thus, four bits could represent the decimal numbers -8 through 7. Table 9 gives examples using a four-bit number. The decimal value is calculated by summing across the value of each column containing a 1.

$-(2^3) = -8$	$2^2 = 4$	$2^1 = 2$	$2^0 = 1$	Decimal Value
0	0	1	1	3
1	1	0	1	-3
0	1	0	1	5

1	0	1	1	-5
---	---	---	---	----

**Table 9: Binary to Decimal Example**

The more often the signal is sampled, the less the difference can be. This allows transmission of fewer bits per sample. As long as the number of bits required per sample reduces in a greater proportion than the frequency of samples required, the required bandwidth reduces. For example, assume that if sampling twice as often, three bits can express the signal difference. The required bandwidth now reduces to 48 kbps.

$$16,000 \text{ samples/second} * 3 \text{ bits/sample} = 48,000 \text{ bits/second}$$

The ITU-T has developed a standard based on ADPCM known as recommendation G.726. The standard allows for bit rates of 40, 32, 24, or 16 kbps, but most implementations use 32 kbps.

### CODE-EXCITED LINEAR PREDICTION (CELP)

Another set of methods for decreasing the required bandwidth for voice is known as Code-Excited Linear Prediction (CELP). The recent improvements in processing power have allowed CELP to compress voice to as low as 4.8 kbps, with minor degradation of voice quality. Current work is being performed to lower this bit rate to 2.4 kbps and below.

CELP takes advantage of known characteristics of voice to “predict” the voice waveforms. The processes and algorithms are somewhat complex and each individual algorithm varies. A general synopsis of a typical algorithm follows.

The CELP algorithm first divides the speech into frames. Each frame is further divided into a number (typically four) of sub-frames. For each sub-frame, the encoder computes a set of filter coefficients for the short-term synthesis filter used to model the characteristics of the speech. The excitation for this filter is determined for each sub-frame, and is given by the sum of scaled entries from two codebooks. The first codebook is adaptive, used to model the long-term periodicity present in the speech. For each sub-frame, an index and a gain—which minimizes the error between the reconstructed and the original speech samples—are computed for this codebook. The second codebook is fixed. It contains pseudo-random codes that correspond to the actual speech sampled. Again, each sample produces a corresponding codebook index and gain. Only the indexes and gains from the two codebooks are transmitted. At the decoder, the scaled entries from the two codebooks pass through a synthesis filter to reproduce the speech. Typically, the speech also passes through a post-filter to improve its perceptible quality.

The ITU-T has developed several standards based on CELP, ranging in bit rates and voice quality. Recommendation G.728 currently is a popular implementation because of the low bit-rate (16 kbps) and low-delay of processing. The new recommendation G.723.1 using Algebraic CELP (ACELP) or multi-pulse, maximum-likelihood quantization (MP-MLQ) can achieve bit rates of 5.3 kbps and 6.3 kbps, respectively. Recommendation G.729, known as constant-structure algebraic CELP (CS-ACELP), is typically used for cellular telephony at 8 kbps and is also one of Cisco’s most popular encoding schemes.

### CONFIGURATION

Selecting an encoding scheme is relatively simple. The configuration requires at the most one command entered in dial-peer configuration mode. The default codec is G.729 at 8

kbps, so it is not often that one deviates from that. Only when bandwidth is unlimited (choose G.711) or precious (choose G.723 if possible) should you not use the default.

From configuration mode, the following command enters dial peer configuration mode.

```
Router(config)#dial-peer voice number {voip, vofr, voatm}
```

To change the codec from dial-peer configuration mode, enter the command below.

```
Router(config-dialpeer)#codec {type}
```

The type of codec can be any of the above listed schemes that your hardware and/or software support.

---

## ENCAPSULATION

Analog voice signals are now in the digital format required for encapsulation. To the network these bits, although originating from voice, are pure data. The network handles this data the same as it handles any other application. The data is encapsulated as it moves down the protocol stack at the sender, and de-encapsulated as it moves up the protocol stack at the destination.

The User Datagram Protocol (UDP)—a low-overhead, connectionless protocol—is used for voice traffic because guaranteed delivery is not required. To provide for timestamping and sequencing, RTP is placed above UDP.

---

## TRANSPORT

Currently, Cisco supports five methods for transporting voice across a data network, including Voice over HDLC (VoHDLC), Voice over IP (VoIP), Voice over Frame Relay (VoFR), Voice over IP over Frame Relay (VoIPoFR), and Voice over ATM (VoATM).

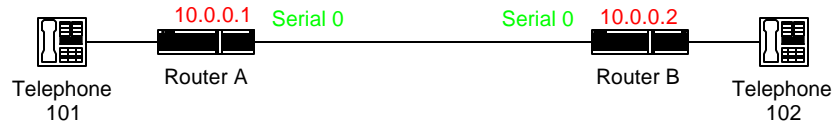
### VOICE OVER HDLC

Voice over HDLC is limited primarily for point-to-point communication. This is typically a major disadvantage. VoHDLC can span multiple hops; however, each site must completely de-compress and re-compress each packet for delivery. This leads to added processor overhead and added delay. VoHDLC also cannot reroute traffic in the event of network outages.

Its major advantage is the small packet overhead required. HDLC requires only five bytes of overhead. Thus, for point-to-point low-bandwidth requirements, VoHDLC is a viable option.

### VOHDLC EXAMPLE

Figure 6 depicts two analog phones connected via FXS ports to the network. The router is the Cisco MC3810 with an FXS port. The two routers connect logically via a direct, HDLC connection. The pertinent configuration is below. Notice that the session-target is the local interface only.



**Figure 6: Simple VoHDLC Example**

```
Router A
!
interface Serial0/0
    ip address 10.0.0.1 255.0.0.0
    voice-encap 512
    clockrate 64000
!
voice-port 1/1
!
dial-peer voice 1 pots
    destination-pattern 101
    port 1/1
!
dial-peer voice 2 vohdlc
    destination-pattern 102
    session target Serial0/0
-----
Router B
!
interface Serial0/0
    ip address 10.0.0.2 255.0.0.0
    voice-encap 512
!
voice-port 1/1
!
dial-peer voice 1 pots
    destination-pattern 102
    port 1/1
!
dial-peer voice 2 vohdlc
    destination-pattern 101
    session target Serial0/0
```

## VOICE OVER IP

The Internet Protocol (IP) is the protocol used by the global Internet for addressing and delivery of data. Because it is the Internet protocol, most public and private local-area networks have also adopted this protocol. This allows the possibility of carrying voice traffic encapsulated within an IP packet completely from source to destination.<sup>3</sup>

---

<sup>3</sup> Does this mean one should immediately begin to use the public Internet as a means for transporting voice traffic? Not yet, but the QoS is improving (this is discussed in more detail later).

VoIP has many advantages. IP, and the corresponding dynamic routing protocols, are very tolerant of network outages. It is also easier to manage. Because IP is the prevalent protocol, network engineers and system administrators feel confident maintaining and managing it.

Recall that voice is first encapsulated within an RTP header and then within an UDP header. IP and the Data Link layer further encapsulate the traffic. This overhead begins to accumulate; especially considering the typical voice data size is only 10-30 bytes. This additional overhead increases bandwidth usage. This is one major disadvantage to VoIP.

## VOIP EXAMPLE

Figure 7 depicts two analog phones connected via FXS ports to the network. The routers can be any router with an FXS port. The two routers connect logically via an IP network. Router A and Router B can be directly connected (i.e. same subnet), or can traverse an IP network as in the diagram. The pertinent configuration is below.



**Figure 7: Simple VoIP Example**

```
router A
!
interface serial 2/0/1
    ip address 10.0.0.1 255.0.0.0
!
voice-port 1/0/0
!
dial-peer voice 1 voip
    destination-pattern 102
    session target ipv4:192.168.0.1
!
dial-peer voice 2 pots
    destination-pattern 101
    port 1/0/0
-----
router B
!
interface serial 3/0/1
    ip address 192.168.0.1 255.255.0.0
!
voice-port 1/0/0
!
dial-peer voice 1 voip
    destination-pattern 101
    session target ipv4:10.0.0.1
!
dial-peer voice 2 pots
    destination-pattern 102
```

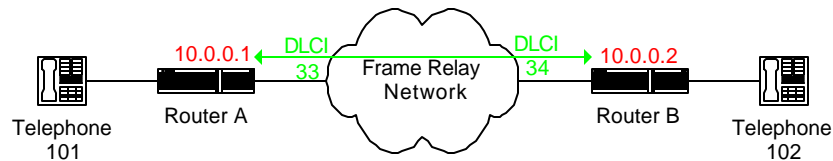
port 1/0/0

## VOICE OVER FRAME RELAY

Frame Relay is a data transportation protocol used to deliver QoS over long distances. Frame Relay networks are private networks—users pay either a flat-fee or a per-use fee. Users pay premiums for a SLA that guarantees QoS—that a guaranteed number of packets from point A will arrive error-free at point B within a certain amount of time. Currently, this makes Frame Relay the most practical method of transporting voice between regional sites.

## VOFR EXAMPLE

Figure 8 depicts two analog phones connected via FXS ports to the network. The routers can be any router with an FXS port. The two routers connect logically via a Frame Relay network. The pertinent configuration is below. Notice that the session-target is the local DLCI, not the far-end DLCI.



**Figure 8: Simple VoFR Example**

```
router A
!
interface serial 2/0/1
    encapsulation frame-relay
    ip address 10.0.0.1 255.0.0.0
    frame-relay interface-dlci 33
!
voice-port 1/0/0
!
dial-peer voice 1 vofr
    destination-pattern 102
    session target serial 2/0/1 33
!
dial-peer voice 2 pots
    destination-pattern 101
    port 1/0/0
-----
router B
!
interface serial 3/0/1
    encapsulation frame-relay
    ip address 10.0.0.2 255.0.0.0
    frame-relay interface-dlci 34
!
voice-port 1/0/0
!
dial-peer voice 1 vofr
```

```

destination-pattern 101
session target serial 3/0/1 34
!
dial-peer voice 2 pots
destination-pattern 102
port 1/0/0
    
```

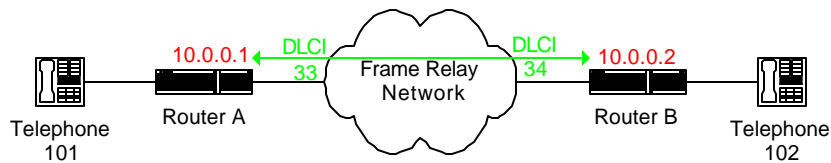
### VOICE OVER IP OVER FRAME RELAY

VoFR uses a Data Link Channel Identifier (DLCI) as the voice session target. Voice over IP over Frame Relay, on the other hand, uses IP as the session target, but the data-link layer protocol just happens to be Frame Relay.

The major advantage of VoIPoFR is the rerouting capability of IP. If the primary link goes down, IP can reroute the call, whether through another Frame Relay DLCI or perhaps through a backup link such as ISDN. With VoFR, if the Frame Relay PVC is down, the call cannot connect. The major disadvantage is the additional overhead caused by IP encapsulation.

### VOIPOFR EXAMPLE

Figure 9 depicts two analog phones connected via FXS ports to the network. The routers can be any router with an FXS port. The two routers connect logically via a Frame Relay network. The pertinent configuration is below. Notice that the session-target is the IP address of the target, not the DLCI of the Frame Relay PVC. This allows the dynamic rerouting of VoIP that VoFR does not allow.



**Figure 9: Simple VoIPoFR Example**

```

router A
!
interface serial 2/0/1
encapsulation frame-relay
ip address 10.0.0.1 255.0.0.0
frame-relay interface-dlci 33
!
voice-port 1/0/0
!
dial-peer voice 1 voip
destination-pattern 102
session target ipv4:10.0.0.2      ← Notice the target is IP!
!
dial-peer voice 2 pots
destination-pattern 101
port 1/0/0
    
```

```

router B
!
interface serial 3/0/1
    encapsulation frame-relay
    ip address 10.0.0.2 255.0.0.0
    frame-relay interface-dlci 34
!
voice-port 1/0/0
!
dial-peer voice 1 voip
    destination-pattern 101
    session target ipv4:10.0.0.1
!
dial-peer voice 2 pots
    destination-pattern 102
    port 1/0/0
    
```

← Notice the target is IP!

## VOICE OVER ATM

Asynchronous Transfer Mode (ATM) is an ITU-T standard designed for multiple service types, such as voice, video, or data. The data is conveyed in small, fixed-size cells that make switching the cells faster—it is more hardware-friendly to switch fixed-sized cells—and guarantee a fixed delay.

VoATM at one time was to be the savior for voice traffic; today few implementations of VoATM exist, mainly due to the assertion of VoFR.

## VoATM EXAMPLE

Figure 10 depicts two analog phones connected via FXS ports to the network. The router is the Cisco MC3810 with an FXS port. The two routers connect logically via an ATM network. The pertinent configuration is below.

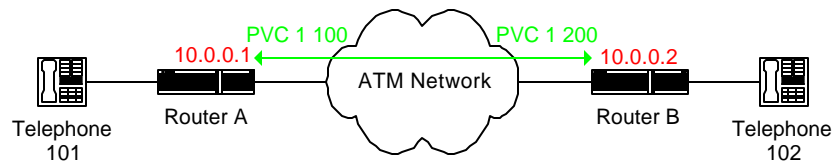


Figure 10: Simple VoATM Example

```

router A
!
controller T1 0
    clock source internal
    mode atm
!
interface atm0
    ip address 10.0.0.1 255.0.0.0
    pvc 1 1 100
        encapsulation aal5mux voice
        vbr-rt 384 192 48
    
```

← PVC For Voice

← **vbr-rt** peak-rate average burst

```
!  
dial-peer voice 1 pots  
  destination-pattern 101  
  port 1/1  
!  
dial-peer voice 202 voatm  
  destination-pattern 102  
  session target ATMO 1  
-----  
router B  
!  
controller T1 0  
  clock source internal  
  mode atm  
!  
interface atm0 point-to-point  
  ip address 192.168.0.1 255.255.0.0  
  pvc 1 1 200  
    encapsulation aal5mux voice  
    vbr-rt 384 192 48  
!  
dial-peer voice 1 pots  
  destination-pattern 102  
  port 1/1  
!  
dial-peer voice 202 voatm  
  destination-pattern 101  
  session target ATMO 1
```

## QUALITY OF SERVICE

QoS simply means having the network provide the necessary resources for applications to perform effectively. The resources can be any network resource, but the two most popular (and most needed) resources are delay and bandwidth.

QoS is not an absolute term. Take delay as an example. For file transfers, the QoS requires packet delivery and an acknowledgment returned before the TCP timers expire. Would you say this network delivers QoS? SNA requires the packet delivery faster because the SNA timers are not as forgiving. Voice is even more intolerant of delay. The moral of the story is that QoS is relative to the application. Thus, one must consider these varying levels of QoS needed when designing and configuring the network.

The sections that follow lay out several means of achieving QoS for your voice applications. First, a delay budget is presented to give the reader an opportunity to see what areas can be optimized. Then, five general methods of achieving voice QoS are presented, including header compression, queuing, packet classification, traffic policing, and traffic shaping.

### DELAY BUDGET

The goal of voice transport, according to ITU-T recommendation G.114, is to keep the end-to-end delay below 150 milliseconds. The voice quality begins to deteriorate rapidly if the delay increases further. A delay budget aids in determining if this goal is achievable and helps determine if and what areas to improve. Table 10 is a sample delay budget. You will need to create your own based upon your coder type, SLAs, configurations, etc.

		Delay Type	Maximum Delay
Look Ahead Delay	Configurable	Constant	10 ms
Coder Delay	Depends on codec	Constant	20 ms
Queuing Delay	Specialized queuing can improve this	Variable	30 ms
Serialization Delay	Depends on line speed, compression does improve	Constant	1.5 ms
Propagation Delay	Can vary depending on network	Variable	30 ms
Dejitter Buffer	Usually configurable	Constant	50 ms
<b>Total</b>			111.5 ms

**Table 10: Sample Delay Budget**

Some definitions and explanations are required to fully clarify the delay budget.

*Look-Ahead Delay*—The time required to collect a certain number of bits for encoding. The formula for look-ahead delay is number of samples collected divided by samples per second. How many data bits should be collected before encapsulation and transportation? Because voice traffic is very sensitive to delay, these bits should be encapsulated and transmitted as quickly as possible. However, this must balance with the utilization of each packet—getting the most data information into each packet to offset the overhead.

*Coding Delay*—The time required to encode the analog voice into digital. Remember, more compression usually means more coding delay. Refer to the previous section Voice-Encoding for a comparison of the different types of encoders and their compression ratios and coding delays.

*Queuing Delay*—The time required for the packet to sit in a queue before transmission onto the wire. This is dependent upon the number of packets awaiting transmission on the interface and the type of queuing configured. The sections that follow demonstrate how to improve this delay.

*Serialization Delay*—The time required to put the frame onto the wire. The formula for this delay is the frame size in bits divided by the line speed in bits per second. Thus increasing the bandwidth of the line will decrease the serialization delay. Compressing either the voice traffic or the packet headers also decreases this delay.

*Propagation Delay*—The time required for the packet to traverse the network from source to destination. This is dependent upon the congestion of the network and the QoS. If you manage the WAN network, you are required to optimize the delay. In most cases, however, a service provider manages the network. Thus, you need only ensure SLAs provide the needed QoS and that the providers are adhering to this SLA.

*Dejitter Buffer*—The amount of time the packet will remain on the destination router before being transformed back into analog voice. This buffer allows for jitter, or variations for the time it may take packets to arrive.

---

## HEADER COMPRESSION

Compressing the header of a VoIP packet has two advantages. One, it decreases the amount of bandwidth needed. Two, it decreases the serialization delay because the packet is now smaller.

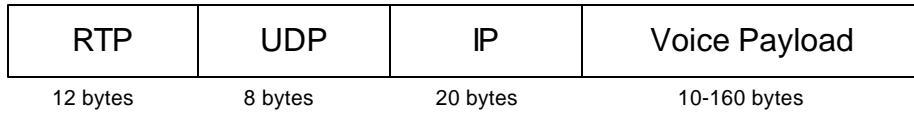
The only disadvantage is the increased processing required compressing the header. For this reason, it is not recommended to enable header compression on links faster than T1 or E1 speeds.

## COMPRESSED RTP (cRTP)

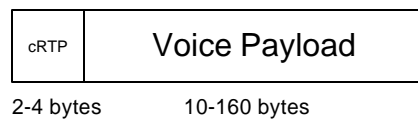
As was stated previously, VoIP uses the Real-Time Transport Protocol. This appends 40-bytes of overhead to the voice payload—20-bytes of IP, 8 bytes of UDP, and 12 bytes of RTP. This becomes significant considering G.729 with one 10-ms sample/frame consists of only 10 bytes per frame. Without compression, that amounts to 80% of each packet being overhead.

cRTP can compress the IP, UDP, and RTP header from 40 bytes to either 2 or 4 bytes, depending on whether a CRC is desired. It is able to do this because several of the fields, such as source/destination address and ports, remain constant throughout the life of the conversation. The changing fields do so at a constant rate, which allows even more compression.

### RTP



### cRTP



**Figure 11: Compressed RTP Comparison**

Each side must periodically transmit the full header to ensure that the other side has the current header. In addition, if something in the header normally constant or changing at a constant rate begins to differ, the full header must be transmitted.

Because cRTP operates on a link-by-link basis, both sides must configure the link RTP header compression. A cRTP passive mode does exist, whereby an interface begins to compress packets only if it receives compressed packets from the other end of the link.

Configuring cRTP on a PPP or HDLC interface is different than configuring it on a Frame Relay interface. The command for implementing it on a PPP or HDLC interface is below. Enter the command from interface configuration mode.

```
Router(config - intf)#ip rtp header-compression [passive]
```

To complicate matters even further, the command for implementing cRTP on a Frame Relay major interface is different from implementing it on Frame Relay sub-interfaces. Table 11 lists both commands.

Command	Comment
frame-relay ip rtp header-compression [passive]	Configure cRTP on Frame Relay major interface. The command applies to all subinterfaces and PVCs under the major interface.
rtp header-compression [active passive]	Configure cRTP on Frame Relay subinterface.

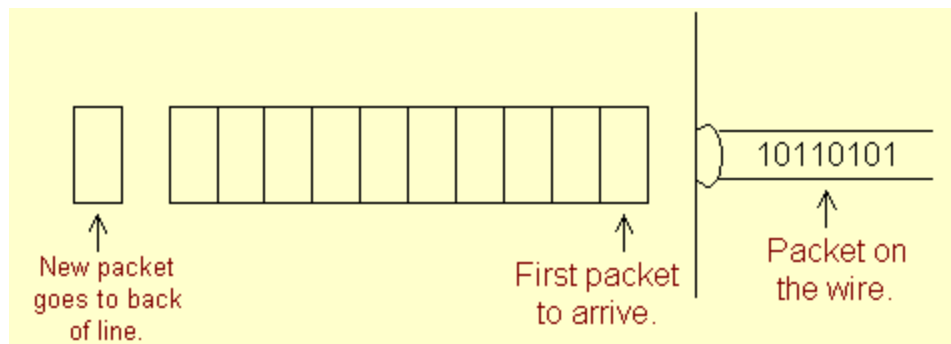
**Table 11: cRTP Frame Relay Commands**

## QUEUEING

A *queue* is the memory buffer that stores packets before transmittal. Multiple queues may exist for a single interface, and each queue may be handled in a different manner. For example, the different queues may be assigned different priorities where the higher priority queue's packets are always transmitted first. *Queueing* is the set of rules that determines the order of queue transmission. Attributes of each packet determine the queue for the packet. Queueing also includes the process of assigning packets to queues.

### FIFO QUEUEING

The simplest type of queueing is first-in first-out (FIFO) queueing. One analogy is a single open lane at a tollbooth. When cars approach, they simply get in the back of the line. The first car in line is the first car serviced, and each car thereafter is serviced in turn. This is the same way FIFO Queueing works. The first packet to arrive is transmitted first, and each packet thereafter is transmitted in turn.



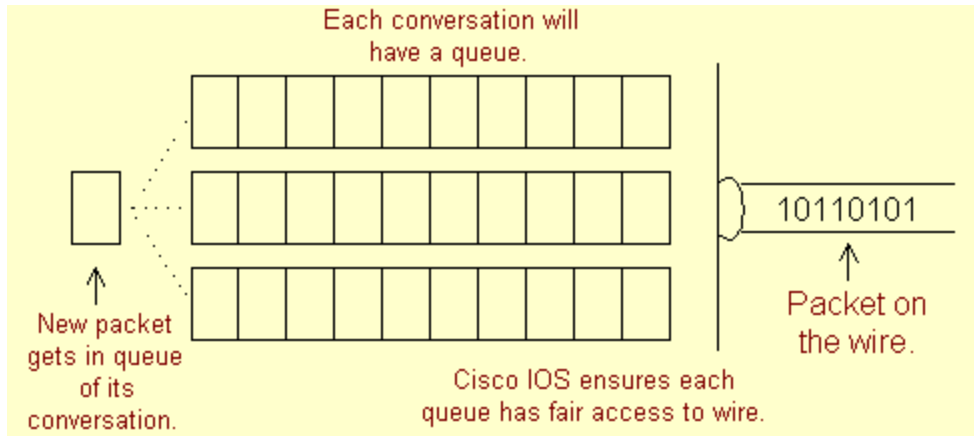
**Figure 12: FIFO Queueing**

Weighted fair queueing (WFQ) is the default queueing enabled on Cisco router interfaces. Thus, to configure FIFO queueing, one must simply disable WFQ. The command below disables WFQ and enables FIFO queueing.

```
Router(config - intf)#no fair-queue
```

### WEIGHTED FAIR QUEUEING

WFQ works by assigning each conversation to a different queue. A source and destination address and a source and destination port define a conversation. WFQ initially may seem to defy logic—it gives low-volume traffic priority over high-volume traffic. Further analysis, however, reveals that low-volume traffic, such as Telnet, requires a faster response time than high-volume traffic, such as file transfers.



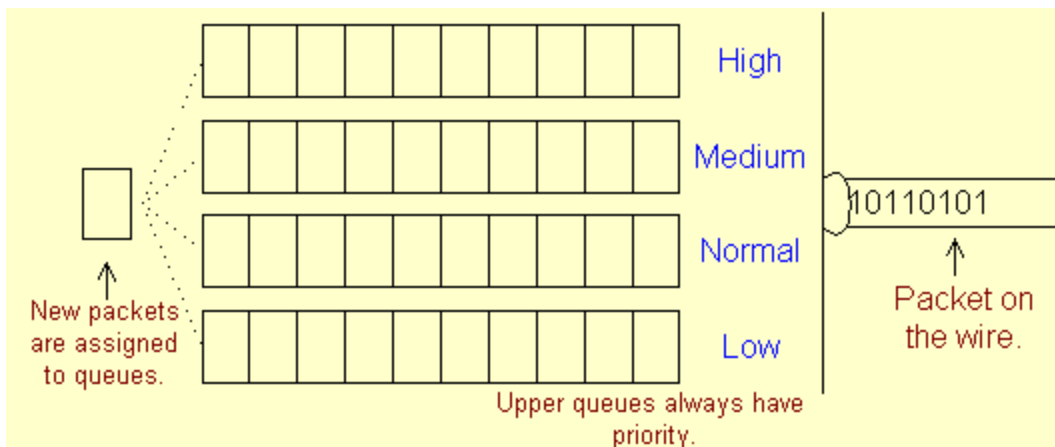
**Figure 13: Weighted Fair Queuing**

Probably the best feature of WFQ is the total automation by the Cisco software. Basic configuration typically requires no configuration at all, because WFQ is the default. If required, the single command for enabling WFQ is below.

```
Router(config - intf)#fair-queue
```

### PRIORITY QUEUING

Priority Queuing systematically transmits queues in a preset order. The system classifies the traffic into one of the four queues—a high, medium, normal, and low queue—based on either protocol type or standard or extended access-lists. The system examines the queues in order, beginning with the high, then moving to the medium, normal, and low queues. If traffic exists in the queue, the system transmits the packets. If no traffic exists in a queue, the system examines the next queue.



**Figure 14: Priority Queuing**

Going back to the tollbooth analogy, priority queuing is similar to the SpeedPass initiated by several states. Those with the special pass need not enter the lines; they have their own lane (queue) that gives them the highest priority.

The advantage of priority queuing is that traffic in the higher queues receives near real-time access to the bandwidth. The disadvantage is that traffic in the higher queues may starve traffic in the lower queues from transmission.

Voice over IP traffic uses destination UDP ports starting at 16384 through  $16384 + 4 * (\text{number of voice ports})$ . If we assume 25 voice ports, VoIP uses ports 16384 through 16484. H.323 uses destination TCP port 1720 for call setup. Therefore, to properly implement priority queuing for voice, the system must give priority to these ports. The extended access-list below gives an example.

```
access-list extended VOICE
  permit udp any any range 16384 16484
  permit tcp any any eq 1720
```

Now assign the extended access-list to the highest queue. The other queues can also be assigned. This example will only classify the voice traffic.

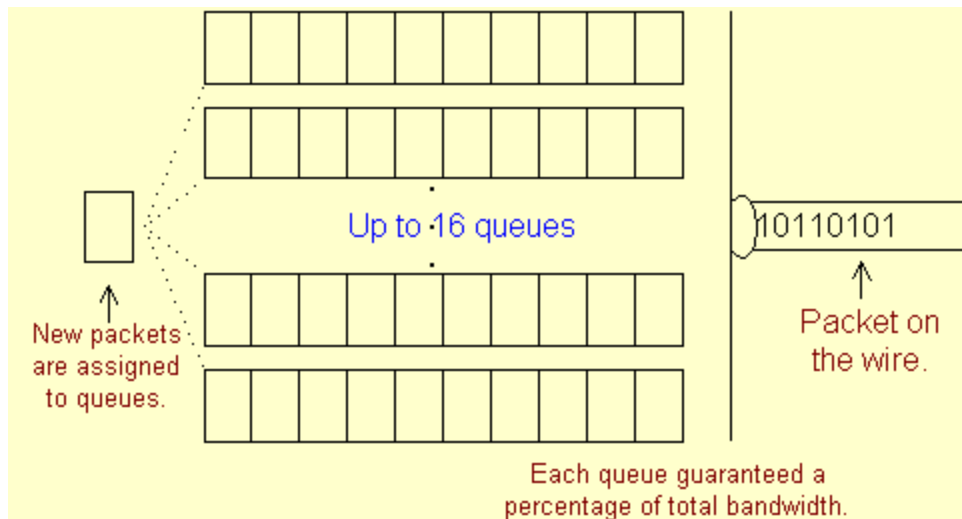
```
priority-list 1 protocol ip high list VOICE
```

The priority queuing now only needs to be assigned to the interface.

```
interface s0
  priority-group 1
```

## CUSTOM QUEUING

Custom queuing assigns a percentage of available bandwidth to up to 16 queues. The router examines the queues in a round-robin fashion. The router transmits a configured number of bytes, and then examines the next queue. This algorithm guarantees a certain percentage of available bandwidth to each queue.



**Figure 15: Custom Queuing**

Custom queuing requires similar configuration to that of priority queuing. First, classify the traffic using an access-list.

```
access-list extended VOICE
  permit udp any any range 16384 16484
  permit tcp any any eq 1720
```

Now assign the traffic to a queue and configure a byte-count. Also, assign a default queue for non-conforming traffic. In the example below, the two equal byte-counts assign half of the available bandwidth to voice and half to the remaining IP protocols.

```
queue-list 1 protocol ip 1 list VOICE
queue-list 1 protocol ip default 2
queue-list 1 queue 1 byte-count 2000
queue-list 1 queue 2 byte-count 2000
```

Lastly, assign the custom queuing to the interface.

```
interface s0
  custom-queue-list 1
```

---

## PACKET CLASSIFICATION

Packet classification is a mechanism that enables the administrator some control over the weighting functions of WFQ. Several mechanisms exist for controlling this weighting, including set the IP precedence bits, Resource Reservation Protocol (RSVP), IP RTP Reserve, and IP RTP Priority. Each is discussed in detail in the following sections.

### IP PRECEDENCE

Setting IP Precedence allows for traffic weighting based on the three bits in the Type of Service (ToS) field of an IP header. The three bits allow for eight (0-7) levels of service. Levels 6 and 7 are reserved for network information, such as keepalives and routing updates. Of the six remaining levels, voice applications should request the highest level of service.

The big advantage of altering the IP Precedence is that no overhead traffic is required for implementation. One disadvantage is that applications can also set the IP Precedence bits. To prevent this, some administrators automatically set these bits to 0 (the lowest precedence) on default gateway routers. This ensures the voice application has the highest precedence.

The simplest way to alter the IP Precedence for voice traffic is via the dial-peer configuration. The command for assigning the IP Precedence value is below.

```
Router(config-dial-peer)#ip precedence value
```

The example below sets the IP Precedence bits to 5 (the highest level for non-router data) for all calls destined to the dial-peer.

```
dial-peer voice 1 vofr
  ip precedence 5
```

Service Type	Level
--------------	-------

Routine	0
Priority	1
Immediate	2
Flash	3
Flash-override	4
Critical	5
Internet	6
Network	7

**Table 12: Type of Service Fields**

One other way to set the precedence of voice traffic is policy routing. The example below sets the IP precedence to critical for voice and to routine for all other applications. See Table 12 for the service types. Do not forget that policy routing must be enabled on the inbound interface.

```
interface s0
  ip policy route-map set-precedence
!
access-list extended VOICE
  permit udp any any range 16384 16484
  permit tcp any any eq 1720
route-map set-precedence permit 10
  match ip address VOICE
  set ip precedence critical
route-map set-precedence permit 20
  set ip precedence routine
```

All traffic that enters via a voice port is considered to originate on that router. For voice traffic to be policy routed that originates on the router, enter a special command from global configuration mode. This command is below.

```
Router(config)#ip policy local route-map route-map-name
```

## RSVP

RSVP allows end-systems to signal, or *reserve*, the required QoS to the network for each specific application. This dynamic nature allows less administration while still maintaining a high level of service. As applications need QoS, they request it and the network responds. If the application terminates or no longer needs the QoS, the application will signal the network to remove the reservation.

RSVP is unique in that the server, as opposed to the client, requests the reservation. This nature allows integration with multicast routing. Multicast routing has one transmitter and multiple receivers. Rather than having the transmitter request QoS for multiple receivers, each receiver can reserve its own required QoS.

Each router in the path from server to client can either accept or deny the request. The router bases this decision on the configuration and the resources available for reservation. The server does receive feedback from each node, so applications that are more intelligent can transmit, choose another route, or choose not to transmit.

By default, RSVP is deactivated on Cisco router interfaces. To use RSVP, enable it on each interface throughout the network supporting voice. Enter the command below to activate RSVP on an interface.

```
Router(config-intf)#ip rsvp bandwidth [interface-kbps] [single-flow-kbps]
```

The interface-kbps is the maximum that RSVP applications can reserve. The single-flow-kbps is the maximum that a single application can reserve. The interface-kbps can never be more than 75% of the interface bandwidth, but a single application can reserve up to the entire amount.

To enable voice applications on Cisco routers to request resources, the dial-peers must be configured to activate the RSVP reservations. The command below, when entered in dial-peer configuration mode, requests an RSVP reservation on behalf of the dial-peer.

```
Router(config – dialpeer)#req-qos [best-effort | controlled-load | guaranteed-delay]
```

The best-effort attribute, which is the default, makes no reservations. This effectively disables RSVP. The controlled-load attribute is the recommended.

The acc-qos command allows monitoring of the QoS levels required. It allows the Cisco router to send a SNMP alert in the event the QoS drops below a specified level. The command below, when entered in dial-peer configuration mode, sets this SNMP trap.

```
Router(config – dialpeer)#acc-qos [best-effort | controlled-load | guaranteed-delay]
```

To display the current RSVP reservations, enter the command below.

```
Router#show ip rsvp reservation [type number]
```

The type and number attributes allow specification of a specific interface.

## IP RTP RESERVE

IP RTP Reserve, as the name implies, allows a static reservation of bandwidth on a particular interface. The system creates a special queue for the reserved traffic, and weights this queue higher than other traffic. This special queue does not necessarily guarantee the voice traffic the prescribed bandwidth; it only assigns a higher weight. The system assigns traffic into this special queue based on UDP port numbers.

Enable the reservation on each interface that the voice traffic flows. To enable IP RTP Reserve, enter the following command from interface configuration mode.

```
Router(config-if)#ip rtp reserve lowest-udp-port range-of-ports maximum-bandwidth
```

The example below configures a static IP RTP reservation of 1 Mbps for voice traffic.

```
interface Ethernet 0/0
```

```
ip rtp reserve 16384 100 1000
```

## IP RTP PRIORITY

IP RTP Priority does guarantee voice traffic a specified amount of bandwidth. IP RTP Priority allows specifying a range of UDP/RTP ports whose voice traffic is guaranteed strict priority service over any other queues or classes. When used in conjunction with WFQ, IP RTP Priority provides strict priority to voice, and the system applies WFQ scheduling to the remaining queues.

One concern for IP RTP Priority is correct allocation of voice bandwidth. During times of congestion, if the voice traffic exceeds the bandwidth specified via IP RTP Priority, the system drops the excess traffic. Thus, it is recommended to allocate some additional bandwidth than actually required for the IP RTP Priority bandwidth. Do not forget to add RTP, UDP, and IP headers into the bandwidth calculation. Layer 2 headers need not be included.

The command below configures IP RTP Priority for a specified port-range for a specified bandwidth on an interface.

```
Router(config-if)#ip rtp priority starting-udp-port-number port-number-range bandwidth
```

The example below configures a static IP RTP priority reservation of 1 Mbps for voice traffic.

```
interface Ethernet 0/0
  ip rtp priority 16384 100 1000
```

---

## TRAFFIC POLICING

The two previous sections provided mechanisms to queue and prioritize traffic. Traffic policing provides a mechanism to regulate or limit traffic. Traffic policing, as opposed to traffic shaping, can actually drop traffic that exceeds the limits set. The section that follows describes Cisco's implementation of traffic policing, known as Committed Access Rate (CAR).

### CAR

CAR is a policing mechanism that allows setting both conform and exceed actions on traffic rates. The CAR-affected traffic can be set by an access-list, so granular control is possible. Often CAR is used at network edges to limit traffic amounts.

CAR does have some major drawbacks. As of now, CAR can only be implemented on interfaces that support Cisco Express Forwarding (CEF). This includes Cisco 7000 series routers with an RSP7000 or Cisco 7500 series routers with at least a VIP2-40. In addition, CAR only works with IP traffic.

Configuration of CAR occurs under interface configuration mode. Apply the rate-limit to either inbound or outbound traffic. The command is below.

```
Router(config-if)#rate-limit {input|output} [access-group [rate-limit] acl-number] bps
burst-normal burst-max conform-action action exceed-action action
```

Table 13 provides a syntax description of the command. Notice that some attributes are listed in bits while others are listed in bytes.

Attribute	Explanation
input output	Applies CAR to inbound or outbound traffic.
access-group	If included, applies CAR only to traffic specified by the access-list.
rate-limit	Specifies that a rate-limit access-list is used. Rate-limit access-lists allow limiting by IP precedence or mac-address.
acl-number	Access-list number
bps	Average rate in bits per second. The value must in increments of 8 kbps.
burst-normal	Normal burst size in bytes. The minimum value is bps divided by 2000.
burst-max	Excess burst size in bytes.
conform-action	Codeword preceding the action to take.
action	Action to take on packets.
exceed-action	Codeword preceding the action to take.

**Table 13: Syntax of rate-limit command**

Table 14 provides a list of the possible actions to take.

Action	Action to take on packet
continue	Evaluate the next rate-limit command.
drop	Drop the packet.
set-prec-transmit new-prec	Set the IP Precedence then transmit the packet.
set-prec-continue new-prec	Set the IP Precedence then continue to the next rate-limit command.
transmit	Transmit the packet.

**Table 14: Rate-limit actions**

One implementation of CAR for voice networks is setting the IP Precedence. The applications on most networks cannot be trusted to set the IP Precedence to routine. Service Providers may also want to limit the amount of traffic received via an access point. The entry point into the network may set the precedence to routine for normal traffic and critical for voice traffic.

Type of Traffic	Bandwidth Allowed	Normal Burst Size	Excess Burst Size	Conform Action	Excess Action
Voice	500 kbps	16000 bytes	24000 bytes	Set IP Precedence to critical	Continue
Remaining Traffic	1.5 Mbps	24000 bytes	32000 bytes	Set IP Precedence to routine	Drop

**Table 15: Example CAR Worksheet**

Table 15 is an example of a worksheet generated from an SLA between an ISP and a customer. The ISP has agreed to guarantee a certain QoS (by setting the IP Precedence) up to 500 kbps of voice traffic. Any additional voice traffic is treated as normal traffic. The ISP has also agreed to accept up to 1.5 Mbps of normal traffic. The example configuration below implements this worksheet for the ISP.

```
interface Hssi0/0/0
    rate-limit input access-group 101 500000000 16000 24000 conform-action
set-prec-transmit 5 exceed-action continue
    rate-limit input 1500000000 32000 48000 conform-action set-prec-transmit 0
exceed-action drop
!
access-list 101 permit udp any range 16384 16484
```

---

## TRAFFIC SHAPING

Traffic Shaping, as the name implies, regulates traffic flow primarily to avoid congestion. Avoiding congestion prevents packet loss. Traffic Shaping works by simply leaving packets in the queue so that only a certain number of packets or bytes are transmitted in a given interval of time.

Traffic shaping could be implemented, for example, when an SLA agreement exists between a customer and a provider and the customer does not want to run the risk of the ISP discarding packets in excess of the maximum agreeable rate.

Two mechanisms exist for traffic shaping on Cisco routers. Generic Traffic Shaping (GTS) and Frame Relay Traffic Shaping (FRTS). Each is discussed in detail in the following sections.

### GTS

Apply GTS to all interfaces types except ISDN, dialup interfaces, and non-GRE encapsulated tunnels. Both interfaces and sub-interfaces support GTS, as does nearly all data-link layer encapsulations. If a packet is queued, GTS uses a WFQ for the delayed traffic.

Much like CAR, GTS requires an average bit rate, burst size, and excess burst size. Unlike CAR, GTS shapes traffic to these rates instead of performing an action such as drop. Table 16 describes each of these attributes.

Attribute	Description
bit-rate	Traffic is shaped to this rate in bits per second. An SLA may specify this.
burst-size	Sustained number of bits transmitted per interval. The interval is the burst-size divided by the bit-rate.
excess-burst-size	Maximum number of bits in the first interval that can exceed the burst-size in a congestion event.

**Table 16: GTS Attributes**

Configure GTS on a per-interface basis. From interface configuration mode, enter the following command to enable GTS.

```
Router(config-if)#traffic-shape rate bit-rate [burst-size [excess-burst-size]]
```

To traffic shape only a subset of traffic based on an access-list, use the traffic-shape group command. Enter the command below from interface configuration mode.

```
Router(config-if)#traffic-shape group access-list-number bit-rate [burst-size [excess-burst-size]]
```

Refer back to Table 15 for an example of a worksheet generated from an SLA between an ISP and a customer. Recall that the ISP has agreed to guarantee a certain QoS (by setting the IP Precedence) up to 500 kbps of voice traffic. The ISP treats any additional voice traffic as normal traffic. The ISP has also agreed to accept up to 1.5 Mbps of normal traffic.

The customer wants to limit the discarding of packets. For this reason, they decide to implement GTS. The configuration below implements GTS based on the worksheet for the customer.

```
interface serial 0
    traffic-shape group 101 500000000 16000 24000
    traffic-shape group 1 1500000000 24000 32000
!
access-list 101 permit udp any range 16384 16484
access-list 101 permit tcp any eq 1720
access-list 1 permit any
```

To verify that traffic shaping is operational, use the command below.

```
Router#show traffic-shape [statistics]
```

## FRTS

FRTS allows traffic shaping implementation on a per-PVC basis. GTS can also do this, but it requires much more configuration. Each subinterface must use only one PVC, and GTS must be enabled on each. With FRTS, each PVC and/or subinterface inherits the traffic shaping attributes of the main interface. Of course the subinterface can implement FRTS to override this.

It is most important to implement Traffic Shaping on Frame Relay links. Because Frame Relay is a data-link layer protocol, it cannot examine the network layer IP Precedence field. Thus, if congestion occurs Frame Relay cannot discriminate between critical and routine traffic.

FRTS also enables dynamically transmitting more or less bandwidth depending on the value of the Backward Explicit Congestion Notification (BECN) and Forward Explicit Congestion Notification (FECN) bits. The BECNs and FECNs simply notify the router that the link is congested. With FRTS, the router “throttles” back the bandwidth.

First enable FRTS on the interface by using the command below.

```
Router(config-if)#frame-relay traffic-shaping
```

Next, enter the attributes of the traffic shaping from map-class configuration mode. Entering this mode requires use of the following command.

```
Router(config)#map-class frame-relay map-class-name
```

Once in map-class configuration mode, the commands listed in Table 17 define the attribute of FRTS.

Command	Description
frame-relay cir [in out] bps	Defines the CIR given by the provider.
frame-relay bc [in out] bits	Defines the committed burst size, $B_c$ , which is also given by the provider.
frame-relay bc [in out] bits	Defines the excess burst size, $B_e$ , which is also given by the provider.
frame-relay adaptive-shaping {becn foresight}	Enables rate adjustment in response to BECN or ForeSight messages.

**Table 17: Frame Relay Map Class Commands**

Now, apply the map-class to an interface. Accomplish this through the command below entered in interface configuration mode.

```
Router(config-if)#map-group map-class-name
```

Frame Relay has only one bit to differentiate QoS. That bit is the discard eligible (DE) bit. If this bit is set, the packet is the first discarded during congestion. Of course, the network may discard any packet if congestion warrants.

To specify which traffic should have the DE bit set, enter the following command in global configuration mode.

```
Router(config)#frame-relay de-list list-number {protocol protocol|interface type number}  
characteristic
```

The protocol can be nearly any Cisco IOS capable protocol. Table 18 lists the possible characteristics.

Characteristic	Description
fragments	Packets with the IP fragmented bit set.
tcp port	Packets to or from a specified TCP port.
udp port	Packets to or from a specified UDP port.
list access-list-number	Packets that meet the access-list.
gt bytes	Packets with a length greater than bytes, including 4 bytes of FR encapsulation.
lt bytes	Packets with a length less than bytes, including 4 bytes of FR encapsulation.

**Table 18: DE-list Characteristics**

To apply a de-list to a given interface and PVC, enter the following command from interface configuration mode.

```
Router(config-if)#frame-relay de-group list-number dcli
```

PVC	Type of Traffic	Bandwidth Allowed (CIR)	Normal Burst Size (B <sub>n</sub> )	Excess Burst Size (B <sub>e</sub> )	Discard Eligible?
100	Voice	500 kbps	144000 bits	192000 bits	No
200	Remaining Traffic	1.5 Mbps	192000 bits	256000 bits	Yes

**Table 19: Example FRTS Worksheet**

Refer to Table 19 for an example of a worksheet generated from an SLA between a Frame Relay provider and a customer. The Frame Relay provider has allotted two PVCs. The voice traverses PVC 100 and the remaining traffic PVC 200. The SLA guarantees a certain QoS up to 500 kbps for PVC 100. The ISP has also agreed to accept up to 1.5 Mbps of normal traffic. As congestion occurs, the normal traffic is discarded first because it has the discard eligible bit set. The network may discard voice traffic as well, depending on the amount of congestion.

The configuration below implements this worksheet for the customer.

```
interface serial 0
    encapsulation frame relay
    frame-relay traffic-shaping
!
interface serial 0.1 point-to-point
```

```
        frame-relay class VOICE
!
interface serial 0.2 point-to-point
    frame-relay interface-dlci 200
    frame-relay de-group 101 200
    frame-relay class OTHER
!
map-class frame-relay VOICE
    frame-relay cir 500000
    frame-relay bc 144000
    frame-relay be 192000
    frame-relay adaptive-shaping becn
!
map-class frame-relay OTHER
    frame-relay cir 1500000
    frame-relay bc 192000
    frame-relay be 256000
    frame-relay adaptive-shaping becn
!
access-list 101 permit udp any range 16384 16484
access-list 101 permit tcp any eq 1720
!
```

---

## SUMMARY

---

Yesterday's circuit-switched networks are becoming obsolete, giving ground to today's packet-switched voice and data networks. The cost-saving advantages are obvious, and applications are being built to enable these networks to also generate revenue. The disadvantages have hampered the implementation of these networks, but engineers are designing faster and more-productive products to overcome these disadvantages.

Analog and digital voice ports enable the data network to interface the telephony equipment. Analog voice ports connect to an analog telephone, the PSTN, or a PBX. Digital voice ports connect to a PBX or the PSTN.

Any good voice over packet network design begins with a dial plan. The dial plan, at the very least, should map telephone numbers to addresses. Other attributes, such as voice encoding scheme and QoS method, should also be included.

H.323 is the primary protocol used for session initiation and call setup and clearing. It certainly is not the best, but because of backward-compatibility reasons, no other protocol has gained market share. Other protocols such SIP and MGCP are gaining ground because they are simpler and more robust.

PCM is the old standard of voice encoding. It is also the fastest, but it requires the most bandwidth. ADPCM is the first generation of voice compression. The second generation is CELP. The golden rule is the more compression involved, the longer it takes to encode the voice, and more processing power is required.

Voice over IP, especially over Frame Relay, is the most popular transport of voice traffic. Voice over Frame Relay is also very popular because it uses less packet overhead. Voice over HDLC and Voice over ATM are sparsely used.

Quality of Service is paramount for high-quality voice. Generating a delay budget greatly assists in determining those areas that can be improved and those that cannot. Compressing of RTP headers greatly reduced the overhead of VoIP traffic. Applying one of the four different types of queuing certainly decreases the amount of time a packet may sit in a router's buffer. An alternative to queuing is packet classification, using IP Precedence, RSVP, IP RTP Reserve, or IP RTP Priority. Traffic policing and traffic shaping limit the traffic transmission rate. Traffic policing discards the excess traffic, while traffic shaping simply buffers the traffic until the congestion ceases.

The paper's primary goal was to prepare CCIE candidates for the voice section of the laboratory exam. Configuring simple Voice over IP (or any other protocol) is a relatively simple task. There are small details that can quickly begin to toughen the task. Add QoS to voice and the tasks becomes even harder. Nevertheless, this paper will certainly give you a head start on attaining that coveted title of CCIE.