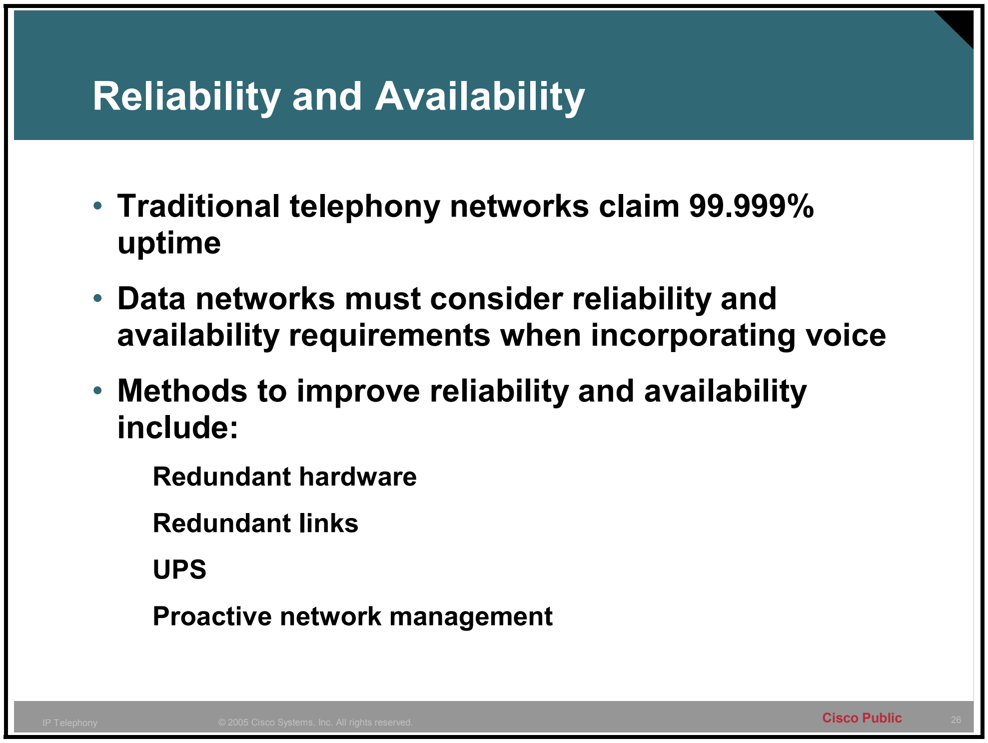
Challenges and Solutions in VoIP

Challenges in VoIP

The traditional telephony network strives to provide 99.99 percent uptime to the user. This corresponds to 5.25 minutes per year of down time. Many data networks cannot make the same claim. This topic describes methods that you can use to improve reliability and availability in data networks.



To provide telephony users the same—or close to the same—level of service as they experience with traditional telephony, the reliability and availability of the data network takes on new importance.

When the data network goes down, it may not come back up for minutes or even hours. This delay is unacceptable for telephony users. Local users, with network equipment such as voice-enabled routers, gateways, or switches for IP Phones, now find that their connectivity is terminated. Administrators must, therefore, provide an uninterruptible power supply (UPS) to these devices *in addition* to providing network availability. Previously, depending on the type of connection the user had, they received their power directly from the telephone company central office (CO) or through a UPS that was connected to their keyswitch or PBX in the event of a power outage. Now the network devices must have protected power to continue to function and provide power to the end devices.

Network reliability comes from incorporating redundancy into the network design. In traditional telephony, switches have multiple redundant connections to other switches. If either a link or a switch becomes unavailable, the telephone company can route the call in different ways. This is why telephone companies can claim a high availability rate.

High availability encompasses many areas of the network. In a fully redundant network, the following components need to be duplicated:

- Servers and call managers
- Access layer devices, such as LAN switches
- Distribution layer devices, such as routers or multilayer switches
- Core layer devices, such as multilayer switches
- Interconnections, such as WAN links, even through different providers
- Power supplies and UPSs

In some data networks, a high level of availability and reliability is not critical enough to warrant financing the hardware and links required to provide complete redundancy. If voice is layered onto the network, these requirements need to be revisited.

With Cisco Architecture for Voice, Video and Integrated Data (AVVID) technology, the use of Cisco CallManager clusters provides a way to design redundant hardware in the event of Cisco CallManager failure. When using gatekeepers, you can configure backup devices as secondary gatekeepers in case the primary gatekeeper fails. You must also revisit the network infrastructure. Redundant devices and Cisco IOS services, like Hot Standby Router Protocol (HSRP), can provide high availability. For proactive network monitoring and trouble reporting, a network management platform such as CiscoWorks2000 provides a high degree of responsiveness to network issues.

Bandwidth Requirements in VoIP

This topic describes the bandwidth that each coder-decoder (codec) uses and illustrates its impact on total bandwidth.

Bandwidth Implications of Codec

Codec	G.711	G.726 r32	G.726 r24	G.726 r16	G.728	G.729	G.723 r63	G.723 r53
Bandwidth	64 kbps	32 kbps	24 kbps	16 kbps	16 kbps	8 kbps	6.3 kbps	5.3 kbps

One of the most important factors for the network administrator to consider while building voice networks is proper capacity planning. Network administrators must understand how much bandwidth is used for each Voice over IP (VoIP) call. With a thorough understanding of VoIP bandwidth, the network administrator can apply capacity-planning tools.

Following is a list of codecs and their associated bandwidth.

- The G.711 pulse code modulation (PCM) coding scheme uses the most bandwidth. It takes samples 8000 times per second, each of which is 8 bits in length, for a total of 64000 bps.
- The G.726 adaptive differential pulse code modulation (ADPCM) coding schemes use somewhat less bandwidth. While each coding scheme takes samples 8000 times per second like PCM, it uses 4, 3, or 2 bits for each sample. The 4, 3, or 2 bits for each sample results in total bandwidths of 32000, 24000, or 16000 bps.
- The G.728 low delay-code excited linear prediction (LD-CELP) coding scheme compresses PCM samples using codebook technology. It uses a total bandwidth of 16000 bps.
- The G.729 and G.729a Conjugate Structure Algebraic Code Excited Linear Prediction (CS-ACELP) coding scheme also compresses PCM using advanced codebook technology. It uses 8000 bps total bandwidth.

- The G.723 and G.723a multipulse maximum likelihood quantization (MPMLQ) coding schemes use a look-ahead algorithm. These compression schemes result in 6300 or 5300 bps.

The network administrator should balance the need for voice quality against the cost of bandwidth in the network when choosing codecs. The higher the codec bandwidth, the higher the cost of each call across the network.

Bandwidth Requirements in VoIP: Impact of Voice Samples

This topic illustrates the effect of voice sample size on bandwidth.

Impact of Voice Samples

Codec	Bandwidth	Sample	Packets
G.711	64000	240	33
G.711	64000	160	50
G.726r32	32000	120	33
G.726r32	32000	80	50
G.726r24	24000	80	25
G.726r24	24000	60	33
G.726r16	16000	80	25
G.726r16	16000	40	50
G.728	16000	80	13
G.728	16000	40	25
G.729	8000	40	25
G.729	8000	20	50
G.723r63	6300	48	16
G.723r63	6300	24	33
G.723r53	5300	40	17
G.723r53	5300	20	33

Voice sample size is a variable that can affect total bandwidth used. A voice sample is defined as the digital output from a codec digital signal processor (DSP) that is encapsulated into a protocol data unit (PDU). Cisco uses DSPs that output samples based on digitization of 10 ms worth of audio. Cisco voice equipment encapsulates 20 ms of audio in each PDU by default, regardless of the codec used. You can apply an optional configuration command to the dial peer to vary the number of samples encapsulated. When you encapsulate more samples per PDU, total bandwidth is reduced. However, encapsulating more samples per PDU comes at the risk of larger PDUs, which can cause variable delay and severe gaps if PDUs are dropped.

Example

Using a simple formula, it is possible for you to determine the number of bytes encapsulated in a PDU based on the codec bandwidth and the sample size (20 ms is default):

Bytes_per_Sample = (*Sample_Size* * *Codec_Bandwidth*) / 8

If we apply G.711 numbers, the formula reveals the following:

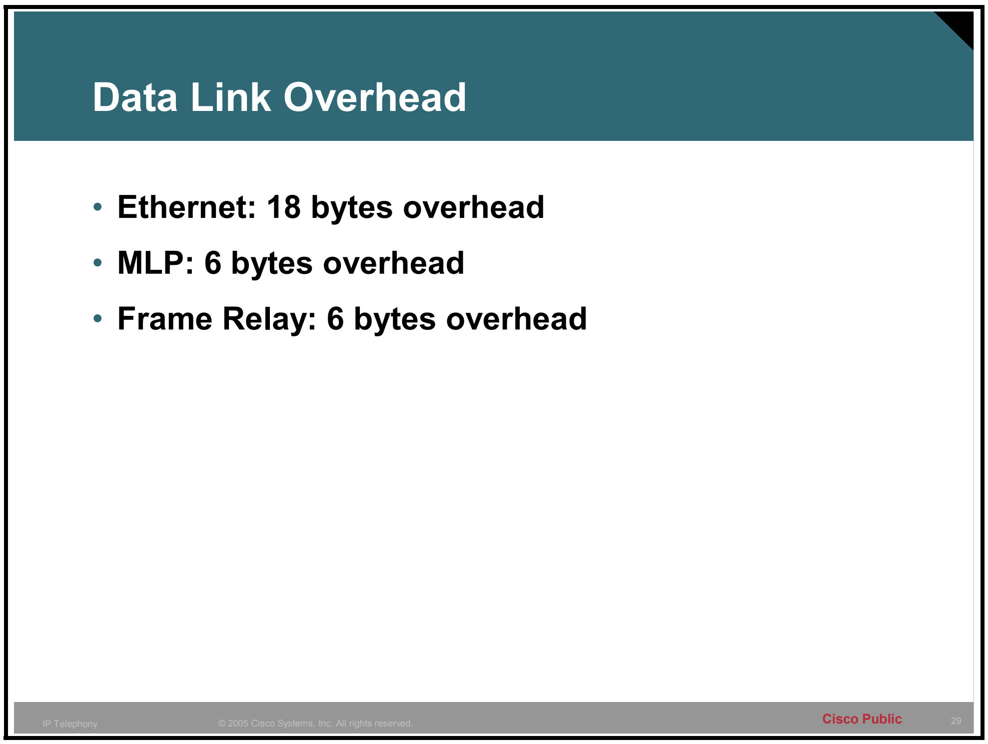
Bytes_per_Sample = (.020 x 64000) / 8

Bytes_per_Sample = 160

The figure illustrates various codecs and sample sizes and the number of packets that are required for VoIP to transmit one second of audio. The larger the sample size, the larger the packet, and the fewer the encapsulated samples that have to be sent (which reduces bandwidth).

Bandwidth Requirements in VoIP: Data Link Overhead

This topic lists overhead sizes for various Layer 2 protocols.



Another contributing factor to bandwidth is the Layer 2 protocol used to transport VoIP. VoIP alone carries a 40-byte IP/User Datagram Protocol/Real-Time Transport Protocol (IP/UDP/RTP) header, assuming uncompressed RTP. Depending on the Layer 2 protocol used, the overhead could grow substantially. The larger the Layer 2 overhead, the more bandwidth required to transport VoIP. The following points illustrate the Layer 2 overhead for various protocols:

- **Ethernet II:** Carries 18 bytes of overhead; 6 bytes for source MAC, 6 bytes for destination MAC, 2 bytes for type, and 4 bytes for cyclic redundancy check (CRC)
- **Multilink Point-to-Point Protocol (MLP):** Carries 6 bytes of overhead; 1 byte for flag, 1 byte for address, 2 bytes for control (or type), and 2 bytes for CRC
- **FRF.12:** Carries 6 bytes of overhead; 2 bytes for data-link connection identifier (DLCI) header, 2 bytes for FRF.12, and 2 bytes for CRC

Bandwidth Requirements in VoIP: Total Bandwidth Required

This topic calculates the total bandwidth required for a VoIP call using codec, data link, and sample size.

Total Bandwidth Required

Codec	Codec Speed	Sample Size	Frame Relay	Frame Relay with CRTP	Ethernet
G.711	64000	240	76267	66133	78933
G.711	64000	160	82400	67200	86400
G.726r32	32000	120	44267	34133	46933
G.726r32	32000	80	50400	35200	54400
G.726r24	24000	80	37800	26400	40800
G.726r24	24000	60	42400	27200	46400
G.726r16	16000	80	25200	17600	27200
G.726r16	16000	40	34400	19200	38400
G.728	16000	80	25200	17600	27200
G.728	16000	40	34400	19200	38400
G.729	8000	40	17200	9600	19200
G.729	8000	20	26400	11200	30400
G.723r63	6300	48	12338	7350	13650
G.723r63	6300	24	18375	8400	21000
G.723r53	5300	40	11395	6360	12720
G.723r53	5300	20	17490	7420	20140

Codec choice, data-link overhead, sample size, and even compressed RTP, all have positive and negative impacts on total bandwidth. To perform the calculations, you must have all of the contributing factors as part of the equation:

- More bandwidth required for the codec = more total bandwidth required
- More overhead associated with the data link = more total bandwidth required
- Larger sample size = less total bandwidth required
- Compressed RTP = significantly reduced total bandwidth required

Example

The following calculation was used to produce the figure:

Total_Bandwidth = ([*Layer_2_Overhead* + *IP_UDP_RTP Overhead* + *Sample_Size*] / *Sample_Size*) * *Codec_Speed*

For example, assume a G.729 codec, 20-byte sample size, using Frame Relay without compressed Real-Time Transport Protocol (cRTP):

Total_Bandwidth = ([6 + 40 +20]/20) * 8000

Total_Bandwidth = 26400 bps

Bandwidth Requirements in VoIP: Effect of VAD

This topic describes the effect of voice activity detection (VAD) on total bandwidth.

Effect of VAD

Codec	Codec Speed	Sample Size	Frame Relay	Frame Relay with VAD
G.711	64000	240	76267	49573
G.711	64000	160	82400	53560
G.726r32	32000	120	44267	28773
G.726r32	32000	80	50400	32760
G.726r24	24000	80	37800	24570
G.726r24	24000	60	42400	27560
G.726r16	16000	80	25200	16380
G.726r16	16000	40	34400	22360
G.728	16000	80	25200	16380
G.728	16000	40	34400	22360
G.729	8000	40	17200	11180
G.729	8000	20	26400	17160
G.723r63	6300	48	12338	8019
G.723r63	6300	24	18375	11944
G.723r53	5300	40	11395	7407
G.723r53	5300	20	17490	11369

IP Telephony Cisco Public 31

On average, an aggregate of 24 calls or more may contain 35 percent silence. With traditional telephony voice networks, all voice calls use 64-kbps fixed-bandwidth links regardless of how much of the conversation is speech and how much is silence. With Cisco VoIP networks, all conversation and silence is packetized. VAD suppresses packets of silence. Instead of sending VoIP packets of silence, VoIP gateways interleave data traffic with VoIP conversations to more effectively use network bandwidth.

VAD provides a maximum of 35 percent bandwidth savings based on an average volume of more than 24 calls.

Note Bandwidth savings of 35 percent is an average figure and does not take into account loud background sounds, differences in languages, and other factors.

The savings are not realized on every individual voice call, or on any specific point measurement.

Note For the purposes of network design and bandwidth engineering, VAD should *not* be taken into account, especially on links that will carry fewer than 24 voice calls simultaneously.

Various features, such as music on hold (MOH) and fax, render VAD ineffective. When the network is engineered for the full voice call bandwidth, all savings provided by VAD are available to data applications.

VAD is enabled by default for all VoIP calls. VAD reduces the silence in VoIP conversations but it also provides comfort noise generation (CNG). Because you can mistake silence for a disconnected call, CNG provides locally generated *white noise* to make the call appear normally connected to both parties.

Example

The figure shows examples of the VAD effect in a Frame Relay VoIP environment. In the example using G.711 with a 160-byte payload, the bandwidth required is 82400 bps. By turning VAD on, you can reduce the bandwidth utilization to 53560bps. This is a savings of 35 percent bandwidth.

