APPENDIX C

IP TELEPHONY

IP Telephony refers to the use of Internet protocols to provide voice, video, and data in one integrated service over LANs, BNs, MANs, and WANs. When most people talk about IP Telephony, they mean *Voice over IP* (*VoIP*). *Voice over ATM* (*VoATM*) and *Voice over Frame Relay* (*VoFR*) are also fairly common, and, as the names suggests, are close cousins to VoIP.

VoIP provides three key benefits compared to traditional voice telephone services. First, it minimizes the need for extra wiring in new buildings (there is one cable for both voice and data, not two). Second, it provides easy movement of telephones and the ability of phone numbers to move with the individual (the number is installed in the telephone, much like an IP address, so anywhere the phone is connected or the phone number programmed, the phone will connect, even if it is halfway around the world). Finally, VoIP is generally cheaper to operate because it requires less network capacity to transmit the same voice telephone call over an increasingly digital telephone network.

VoIP requires a VoIP *Private Branch Exchange* (*PBX*), which connects the organization's internal telephone network into the *public switched telephone network* (*PSTN*). The PBX can also be connected into the organization's WAN and be configured to route calls through the WAN to other organization locations (or even over the Internet), thereby bypassing any long-distance charges in the PSTN (although the call does use up network capacity on the organization's WAN or Internet connection).

The VoIP PBX can be considered a gateway that connects the internal IP telephones to the PSTN.[1] When the IP telephone user lifts the receiver off the hook to place a call, the IP telephone sends a message to the PBX, which responds by sending a dial tone. Once the number has been dialed, the IP telephone then sends a message to the PBX with the telephone number, and the PBX connects the telephone into the PSTN or over the organization's WAN to the VoIP PBX at the other location.

VoIP often uses H.323 at the application layer (see Chapter 2), although *Session Initiation Protocol* (*SIP*) is also common. *Media Gateway Control Protocol* (*MGCP*) and *Skinny Call Control Protocol* (*SCCP*) are other competing application layer protocols although they offer fewer features than H.323 and SIP. All four protocols (H.323, SIP, MCGP, and SCCP) operate at the application layer and contain all the functions needed to start and end telephone calls, as well as to transmit the call data. MCGP and SCCP require

[1]Another type of VoIP PBX uses regular telephones and regular analog circuits inside the organization, and the PBX does all the conversion between the analog telephone and the digital network. This is an older, temporary approach that is quickly disappearing.

the PBX to act as a server and communicate with other MCGP and SCCP devices only through the PBX server; H.323 and SIP can both communicate with other H.323 and SIP clients without needing to go through the PBX (except for telephone number resolution, which is a process much like using a DNS to resolve an application layer name into an IP address [see Chapter 5]). SIP is a newer and more efficient protocol than H.323, and was developed using HTTP as its core, which means that it is simpler to debug and can be easily integrated into Web-based applications and SMTP e-mail applications.

In order to use VoIP, a device such as an IP telephone must support H.323, SIP, MCGP, or SCCP, and also contain a *CODEC* to convert the incoming analog voice signal into a digital bit stream (see Chapter 3). The CODEC is also used at the receiving end to convert the digital data back into the analog voice data. The most commonly used digital voice protocols are 64 Kbps *PCM*, 32 Kbps *ADPCM* (see Chapter 3), and more recent variants on them, such as 8 Kbps G.729 (also called CELP) or 6.3 Kbps G.723 (also called MPMLQ). As might be expected from their lower bandwidth requirements, sound quality can become an issue.

A technique called *Mean Opinion Score (MOS)* has been developed by ITU-T to subjectively rate the voice quality of different CODEC standards. A MOS of 5 is the theoretical maximum (meaning excellent) while a 1 is the lowest score (very poor quality). A 4 is generally regarded as an acceptable level of quality, with barely perceptible levels of quantizing error. PCM has a MOS of 4.1; ADPCM, 3.85; G.729, 3.92; and G.723 (at 6.3 Kbps), 3.9.

Once the CODEC has produced the digital data, the data is surrounded by an application layer packet. H.323 often uses *RTP* (see Chapter 5) for transmission through an IP-based network. The RTP packet in turn is surrounded by UDP, IP, and data link layer packets (e.g., Ethernet, frame relay) for transmission through the network. Voice data packets tend to be very small, and thus the packets added at layers 5, 4, 3, and 2 can add considerable overhead to the transmission. A new version of RTP called *Compressed Real Time Protocol* (CRTP) has been developed that enables the set of RTP, UDP, and IP packets to be compressed to 2 bytes, thus significantly reducing the overhead.

Voice Activity Detection (*VAD*) is another way of reducing the network capacity required to send VoIP calls. With VAD, the end VoIP device monitors the analog voice signal and if the signal drops below a certain amplitude, then the device assumes that no one is speaking and stops sending packets. Since most conversations are silent at least half the time, this can significantly reduce the bandwidth required. VAD must be done carefully to avoiding clipping the speech (cutting off the beginning and end).

The other problem with VAD is that because no data packets are transmitted, the line is completely silent. We are so used to experiencing background noise from interference on traditional analog telephone calls that silence is usually interpreted as meaning the call has been disconnected. Therefore, several vendors have begun to add in background *comfort noise* (sometimes call pink noise) when no data packets are flowing. The VAD device at the sender's location initially sends some ambient background noise packets, which are recognized by the receiving device. After a few seconds, the VAD sending device begins operating normally and no longer sends background noise packets when speech ends. When the receiver detects the slowdown in packets, it repeatedly reproduces the ambient background noise packets to fill the "dead air" so that the receiver does not think the call has been disconnected.

As a result of VAD, the network capacity requirements for VoIP on Ethernet networks ranges from about 8 Kbps using G.729 and CRTP to 45 Kbps using PCM with RTP. On ATM networks, it ranges from about 6 Kbps using G.723 and CRTP to 53 Kbps using PCM.

One of the key issues in VoIP is ensuring that the network has sufficient capacity to send the VoIP packets. While users will accept a few second delays in Web traffic or e-mail, most people cannot tolerate delays in voice conversations over about 250 milliseconds (a quarter of a second). Therefore, VoIP is only practical in networks that enable *Quality of Service* (*QoS*) routing at the IP layer (see Chapter 5) and ideally matching QoS at the data link layer. Most organizations that deploy VoIP, therefore, use IP with QoS and also have policy-based VLANs at the data link layer (see Chapter 8).