IP Telephone Design and Implementation Issues

By William E. Witowsky, Senior Vice President Engineering & Chief Technical Officer,
Telogy Networks, Inc.

Abstract

The growing excitement surrounding the transport of telephony services over traditional data networks such as the Internet, corporate-enterprise intranets and new service provider extranets has led to the development of cost efficient gateway equipment based on embedded systems that converts analog telephony information such as voice and fax into packet data suitable for transport over IP, Frame Relay and ATM networks. As a result, the long-time promise of being able to replace or enhance the traditional PBX by combining voice and data services onto a single network can now finally be realized. In order to do so, a very low-cost telephony device capable of directly exchanging IP packets with the data network is required. Development of this 'IP Telephone' will require the development of a 'system on a chip' which combines digital signal processing functions, microcontroller functions, analog interface, telephone user interface, network interface, and associated glue logic. This article looks at the functional requirements and design of an IP Telephone and examines the implementation issues that must be considered.

1. Introduction

An IP Telephone is a telephone device that transports voice over a network using data packets instead of circuit switched connections over voice only networks. IP Telephony refers to the transfer of voice over the Internet Protocol (IP) of the TCP/IP protocol suite. Other Voice Over Packet (VOP) standards exist for Frame Relay and ATM networks but many people use the terms Voice over IP (VoIP) or “IP Telephony” to mean voice over any packet network.

IP Telephones originally existed in the form of client software running on multimedia PCs for low-cost PC-to-PC communications over the Internet. Quality of Service (QOS) problems associated with the Internet and the PC platform itself resulted in poor voice quality due to excessive delay, variable delay, and network congestion resulting in lost packets, thus relegating VoIP primarily to hobby status. The QOS provided by the Internet continues to improve as the infrastructure is augmented with faster backbone links and switches to avoid congestion, higher access connections to the end user such as xDSL cut down latency, and new protocols like RSVP and techniques like tag switching give priority to delay sensitive data such as voice and video.

Most of the focus on VoIP is currently centered on two key applications. The first is private business network applications. Businesses with remotely located branch offices which are already connected together via a corporate intranet for data services can take advantage of the existing intranet by adding voice and fax services using VoIP technologies. Businesses are driving the demand for VoIP solutions primarily because of the incredible cost savings that can be realized by reducing the operating costs of managing one network for both voice and data and by avoiding access charges and settlement fees, which are particularly expensive for corporations with multi-international sites. Managed corporate intranets do not have the QOS issues which currently plague the Internet; thus voice quality approaches toll quality.

The second key application is VoIP over public networks. This application involves the use of voice gateway devices designed to carry voice to Internet Service Providers, now known as Internet Telephony Service Providers, or to the emerging Next Generation Carriers such as QWEST and Level 3, which are developing IP networks specifically to carry multimedia traffic such as VoIP. ISPs are interested in VoIP as a way of offering new value-added services to increase their revenue stream and break out of the low monthly fixed fee structure currently in place for data services. VoIP also allows them to improve their network utilization. These new services include voice and fax on a per-minute usage basis at rates significantly less than prevailing voice and fax rates for service through the PSTN. The sustainability of this price advantage may be short term, and is dependent on whether the FCC and foreign regulatory agencies will require ISPs to pay the same access charges and settlement fees PSTN carriers are obligated to pay. New carriers such as Qwest and Level 3 are interested in VoIP because data networks are more efficient than traditional voice networks. In the near term, these new carriers can avoid the access charges and settlement fees which account for up to 42% of the cost of a long distance call. In the long term, IP networks are more efficient for a wide range of new applications, particularly multimedia applications enabling convergence of voice, video, data, and fax. For example, web-enabled call centers, a new application powered by IP networks, will greatly enhance the ability of companies to deliver world class customer service. Carriers, too, are interested in VoIP, primarily for competitive reasons. Although VoIP will cannibalize some of their POTS services, they have wisely determined that they too must compete in this rapidly growing marketplace. The market projections are too staggering to be ignored: according to IDC, 10% of the world's fax market could be on the internet in 2 - 3 years, and by 2002, the Internet could carry 11% of US and international long distance traffic.

Technology advances that enable the rapid deployment of Voice Over Packet (VOP) solutions include the advent of low-cost, low-power and high performance digital signal processors (DSPs) and RISC cores to perform all of the CPU-intensive conversion functions for packetizing voice and fax. Also, the arrival of industry standards for voice over packet will for the first time allow interoperability of devices from different manufacturers. Recent standards include ITU H.323 (voice) and T.38 (facsimile) for Voice over IP (VoIP), Frame Relay Forum FRF.11 for voice/fax over Frame Relay networks and ATM Forum Voice Transport over ATM (VToA).

Figure 1 shows examples of IP Telephone applications. On the corporate enterprise, IP Telephones connect via the Ethernet LAN along with PCs. These telephones interact with a VOP-enabled PBX for call setup and administration and access to the PSTN and/or external packet network.

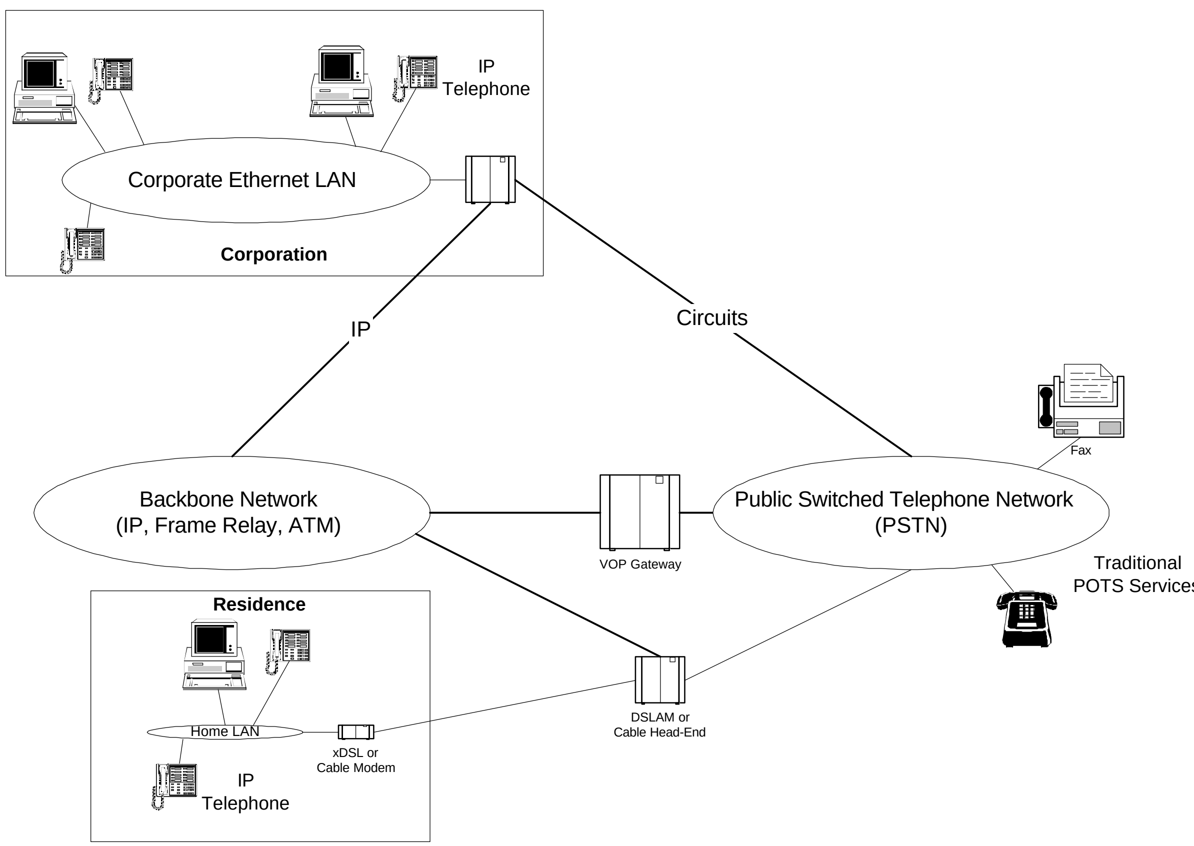

Figure 1 IP Telephone Application

IP Telephones will appear initially in the business environment as a low-cost solution for smaller businesses that would otherwise require a key system or low-end PBX. The advantage of an IP Telephone include having one wiring system for both voice and data, better scalability as additional stations are added to the system, and the ability to mix and match IP Telephones from different manufacturers.

IP Telephones have several advantages over using multimedia PCs with client software: lower latency due to an embedded system implementation, familiar user paradigm of using a phone vs. a "PC-enabled phone", greater reliability, and lower station cost where a PC is also not required, e.g., conference room, production floor, etc.

Note that when considering IP telephones for home use, the network interface available today is typically a dial-up modem connection (V.34bis, V.90) supporting data rates on the order of 30-40kbps. As such, most home users will access voice over packet networks via a local PSTN call to a VOP gateway residing in an ITSP Point of Presence (POP) using a standard POTS phone. In the not too distant future, with the deployment of cable modems and xDSL services providing permanent, high-speed IP connections to the home, there will be a strong demand for IP telephones in residences as well.

Typically these IP telephones will connect to a cable modem or xDSL modem via a high-speed interface such as Ethernet or Universal Serial Bus (USB). There are also emerging home communications standards such as being presented by the Home Phoneline Networking Alliance (www.phonelan.org) for providing LAN capabilities using existing home phone wiring and Home RF (www.homerf.org) which provides wireless communications within the home. In this new residential environment, IP Telephones will attach to the home LAN and have access to the data network and the PTSN via either an xDSL or cable modem which communicates to DSLAM or cable system head-end equipment (refer to Figure 1).

2. Reference Design

Figure 2 below shows a block diagram of a reference design of an IP Telephone. An IP Telephone consists of the following components: User Interface, Voice Interface, Network Interface, and Processor Core and associated logic.

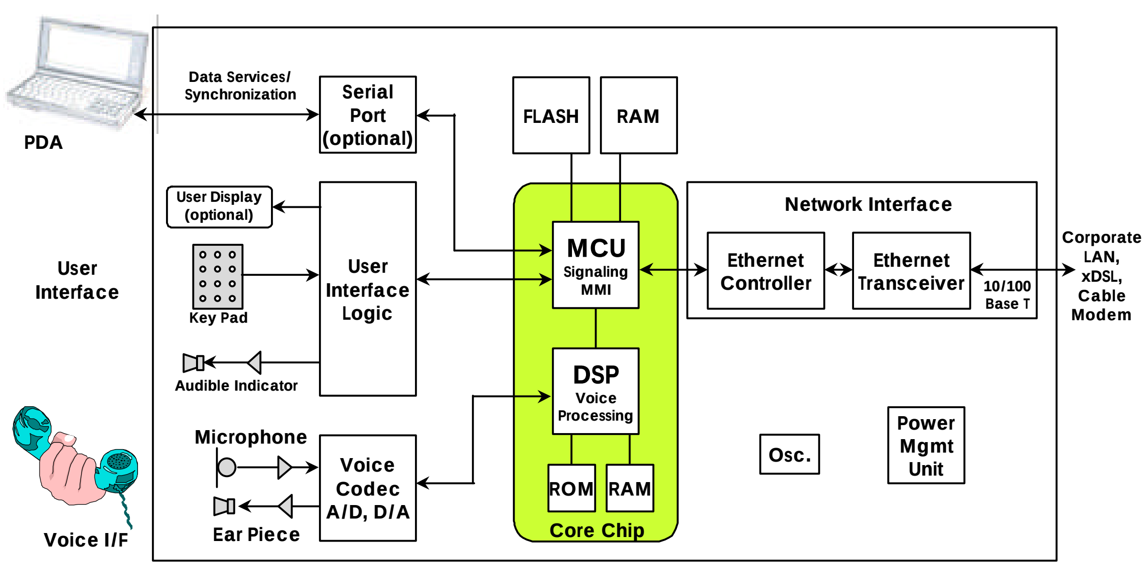

Figure 2 IP Telephone Reference Design

The User Interface provides the traditional user interface functions of a telephone. At a minimum, this consists of a keypad for dialing numbers (0-9, *, #) and an audible indicator for announcing incoming calls to the user. On more sophisticated telephone sets, additional keys are provided for features such as mute, redial, hold, transfer, conferencing, etc. A display is also typically provided for displaying user prompts, number dialed, CallerID information for incoming calls, etc. In certain models, the telephone will be equipped with a serial interface to allow communications to a device such as a PDA to allow synchronization of phone information, facilitate automatic dialing, etc.

The Voice Interface provides the conversion of analog voice into digital samples. Speech signals from the microphone are sampled at a rate of 8 KHz to create a digitized 64kbps data stream to the processor via a pulse code modulation (PCM) codec. Similarly, the processor passes a 64kbps data stream in the return path to the speaker through the PCM codec to convert digital samples back into speech.

The Network Interface allows transmission and reception of voice packets from/to the telephone. For corporate LANs this is most often either 10BaseT or 100BaseT Ethernet running TCP/IP protocols. The IP Telephone may offer a second RJ-45 Ethernet connector to allow a PC to plug in and share one connection to the wall jack.

The Processor Core performs the voice processing, call processing, protocol processing, and network management software functions of the telephone. As shown in Figure 2, this may consist of a Digital Signal Processor (DSP) for the voice-related functions and a Micro Controller Unit (MCU) for the remaining functions. To ensure software upgradeability the telephone will make use of Flash memory.

3. Software Architecture

Figure 3 shows the software architecture of an IP Telephone based on the ITU H.323 standard for VoIP. The software consists of the following major subsystems: User Interface, Digital Signaling Processing, Telephony Signaling Gateway, Network Interface Protocols, Network Management Agent, and System Services. These subsystems are described below.

User Interface

The User Interface subsystem provides the software components that handle the interface to the user of the IP Telephone and consists of the following software modules:

Display Driver
Controls the hardware that generates characters to the display. The Display device itself may range from a simple one line display for CallerID and numbered dialed to a multi-line display including graphic characters.

Keypad Driver
Performs keypad scanning and debounces key presses entered by the user.

Audible Driver
Performs control of the hardware that generates ringing to the user.

User Procedures
Controls the information displayed by the Display Driver and processes user key inputs and converts them into primitives for Call Processing.

Voice Processing

The Voice Processing software is composed of the following software modules:

PCM Interface Unit
Receives PCM samples from the analog interface and forwards then to the appropriate DSP software module for processing. It also forwards processed PCM samples to the analog interface.

Tone Generator
Generates call progress tones to the user and generates in-band DTMF digits to the network based on key presses relayed from the User Interface. For certain, voice codecs, the compression algorithm does not permit faithful transmission of DTMF tones. For those algorithms, e.g., G.723.1, the software generates an in-band message to the network that is used by the remote IP telephone (or gateway) to regenerate the DTMF tone.

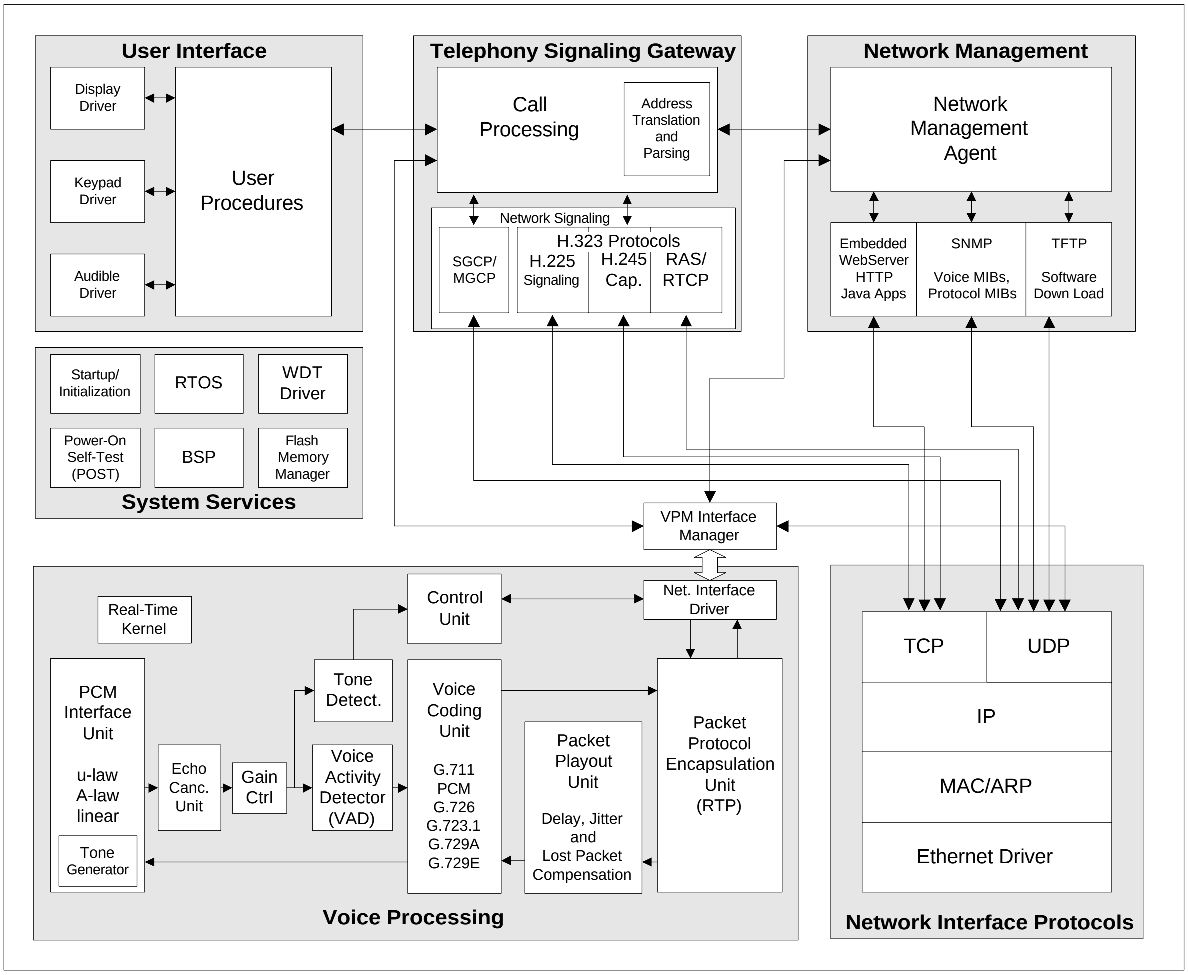

Figure 3 IP Telephone Software Architecture

Line Echo Canceller Unit
Performs ITU G.168 compliant echo cancellation on sampled, full-duplex voice port signals. Echo in a telephone network is caused by signal reflections generated by the hybrid circuit that converts between a 4-wire circuit (a separate transmit and receive pair) and a 2-wire circuit (a single transmit and receive pair). These reflections of the speaker's voice are heard in the speaker's ear. Echo is present even in a conventional circuit switched telephone network. However, it is acceptable because the round trip delays through the network are smaller than 50 msec. and the echo is masked by the normal side tone every telephone generates. Echo becomes a problem in Voice over Packet networks because the round trip delay through the network is almost always greater than 50 msec. Thus, echo cancellation techniques are required. ITU standard G.168 defines performance requirements that are currently required for echo cancellers. Echo is cancelled toward the packet network from the telephone network. The echo canceller compares the voice data received from the packet network with voice data being transmitted to the packet network. The echo from the telephone network hybrid is removed by a digital filter on the transmit path into the packet network.

Acoustic Echo Canceller (optional)
For phone sets featuring speakerphone operation, an acoustic echo canceller is also needed to cancel out echo picked up by the microphone of the received speech. The source of this echo is the reflections of the speaker’s voice off the walls, windows, furniture, etc. in the room where the speakerphone is located.

Voice Activity Detector (VAD)
Detects voice activity and activates or deactivates the transmission of packets in order to optimize bandwidth. When activity is not detected, the encoder output will not be transported across the network. This software also measures Idle Noise characteristics of the interface and reports this information to the Packet Voice Protocol for periodic forwarding to the remote IP Telephone or gateway. Idle noise is reproduced by the remote end when there is no voice activity so that the remote user does will not feel that the line when "dead".

Voice Codec Unit
Performs packetization of the 64 kbps data stream received from the user. Various compression algorithms exist which have different performance characteristics: G.711 PCM which operates at 64 kbps (no compression), G.726 ADPCM which operates at 16, 24, 32 and 40 kbps, G.723.1 which operates at 5.3 or 6.3 kbps and G.729 which operates at 8 kbps. Typically, voice algorithms that perform greater compression require much more processing power. It should be noted that high fidelity audio quality compression algorithms can also be used since an IP Telephone is not subject to the 4 kHz bandwidth restrictions found in the PSTN. This would provide better sounding audio than PCM and allow music to be faithfully reproduced.

Packet Playout Unit
Performs compensation for network delay, network jitter and dropped packets. Many proprietary techniques are used to address these problems since there are currently no standards in place for packet playout.

Packet Protocol Encapsulation Unit
Performs encapsulation of the packet voice data destined for the network interface. For VoIP this encapsulation is per the Real-time Transport Protocol (RTP) which runs directly on top of UDP.

Voice Encryption
Provides optional encryption of the voice packet data prior to transmission over the network to ensure privacy.

Control Unit
Coordinates the exchange of monitor and control information between the Voice Processing module and the Telephony Signaling and Network Management modules. The information exchanged includes software downline load, configuration data, signaling information, and status reporting.

Telephony Signaling Gateway
The Telephony Signaling Gateway (TSG) subsystem performs the functions for establishing, maintaining and terminating a call. The TSG consists of the following software modules:

Call Processing
Performs the state machine processing for call establishment, call maintenance and call tear down.

Address Translation and Parsing
Performs digit collection and parsing to determine when a complete number has been dialed and makes this dialed number available for address translation.

Network Signaling
Performs signaling functions for establishment, maintenance and termination of calls over the IP network. Two standards exist: H.323 and SGCP/MGCP.

H.323 Protocols
H.323 is an ITU standard that describes how multimedia communications occur between user terminals, network equipment, and assorted services on Local and Wide Area IP networks. The following H.323 standards are used for VoIP in an IP Telephone:

- **H.225** – Call Signaling Protocols. Performs signaling for establishment and termination of call connections based on Q.931.

- **H.245 -** Control Protocol. Provides capability negotiation between the two end-points such as voice compression algorithm to use, conferencing requests, etc.

- **RAS** – Registration, Admission, and Status (RAS) Protocol. Used to convey the registration, admissions, bandwidth change, and status messages between IP Telephone devices and servers called Gatekeepers which provide address translation and access control to devices.

- **RTCP –** Real-time Transport Control Protocol (RTCP). Provides statistics information for monitoring the quality of service of the voice call.

SGCP/MGCP Protocols
Simple Gateway Control Protocol (SGCP) is a standard that describes a master/slave protocol for establishing VoIP calls. The slave side or client resides in the gateway (IP telephone) and the master side resides in an entity referred to as a Call Agent. SGCP has been adopted by the Cable Modem industry as part of the DOCSIS standard. SGCP is evolving to the Multimedia Gateway Control Protocol (MGCP).

Network Management

The Network Management subsystem supports remote administration of the IP Telephone by a Network Management System. The Network Management Agent consists of the following software modules:

Network Management Agent
Performs the network management functions of the IP Telephone, including status monitoring and alarm reporting, gathering of statistics in response to SNMP queries, etc. from a Network Management System.

Embedded Web Server (optional)
Provides administration support via a standard web browser. Presents the user with web pages for configuring the IP Telephone and gathering statistics information. May provide Java applets for loading to the user's PC, e.g., for status polling.

SNMP
Performs the Simple Network Management Protocol (SNMP) functions for processing Management Information Base (MIB) Gets and Sets and generation of Alarm Traps.

TFTP
Trivial File Transport Protocol (TFTP) is used to download software updates into Flash memory.

The Network Interface Protocols support communications over the Local Area Network (LAN) and consists of the following software modules:

TCP
The Transport Control Protocol (TCP) provides reliable transport of data including retransmission and flow control. It is used for web queries, and call signaling functions.

UDP
The User Datagram Protocol (UDP) provides efficient but unreliable transport of data. It is used for the transport of real-time voice data since retransmission of real-time data would add too much delay to the voice conversation and be unacceptable. UDP is also used for SNMP and TFTP network management traffic.

IP
The Internet Protocol (IP) provides a standard encapsulation of data for transmission over the network. It contains a source and destination address used for routing.

MAC/ARP
Performs Media Access Control (MAC) management functions and handles Address Resolution Protocol (ARP) for the device.

Ethernet Driver
Configures and controls the ethernet controller hardware, including setting up DMA operations.

System Services
System Services consists of the following software modules:

Startup/Initialization
Provides startup and initialization of the hardware and software components of the IP Telephone.

POST
Provides Power-On Self-Test (POST) functions of the IP Telephone.

RTOS
The Real-Time Operating System (RTOS) provides functions such as task management, memory management, and task synchronization.

BSP
Board Support Package (BSP) software provides hardware interface drivers, interrupt vectors, etc. that allows the real-time operating system to operate on the target hardware platform.

Watch Dog Timer Driver
Provides control of a hardware watchdog timer (WDT) as a control mechanism to prevent the IP Telephone from locking up due to a software or intermittent hardware failure.

Flash Memory Manager
Provides functions for reading/write data from/to the Flash memory.

DSP Interface Manager
Provides the driver for exchanging information between the MCU and DSP, including software download, voice packets and network management functions.

4. Implementation Issues
The following are implementation issues relating to IP Telephones:

Cost
As shown in the previous section, it can be seen that the implementation of an H.323 compliant IP Telephone requires a great deal of voice processing software, network protocol software and network management software. This requires significant processing power and memory which adds cost to the phone. In order reduce the cost of the telephone, many of the functional blocks shown in Figure 2 will typically be integrated into a single chip, e.g., DSP, MCU, Ethernet Controller, PCM codec and associated “glue” logic.

Compression Algorithm
For traffic over switched ethernet LANs where the QOS is excellent and there is plenty of bandwidth for voice and data functions, 64 kbps G.711 PCM voice coding can be used instead of G.723.1 or G.729A to reduce the processing requirements of IP Telephone. The PBX (or Ethernet Hub) can provide compression if required. This reduces the amount of DSP processing resources required and would allow the use of a lower cost DSP or possibly the use of a single RISC processor core for the entire IP Telephone, replacing the DSP/MCU combo. This RISC processor would perform Host Signaling Processing (HSP) functions to handle basic voice processing requirements in addition to its other processing requirements. A DSP may still be needed for more intense processing functions such as acoustic echo cancellation or to provide support of multiple lines and conferencing features.

Legacy Support for 2-wire Analog POTS Devices
There will still be a need to support existing 2-wire analog devices such as Group 3 fax machines. This will require the development of fax adapters which will be a small device that has an RJ-11 connector to emulate a Central Office (CO) interface to the fax machine and an RJ-45 connector to interface to an Ethernet. The adapter will provide loop current, detect off-hook for outgoing faxes and generate ring voltage for incoming faxes. It will demodulate the fax signals (V.21, V.27ter, V.29, and V.17) and convert the fax signals into packet data for transport over the network in a fashion similar to voice.

New Signaling Protocol Standards
The H.323 protocol suite is fairly complex for call establishment. Alternative standards such as Session Initiation Protocol (SIP), the Simple Gateway Control Protocol (SGCP) and Multimedia Gateway Control Protocol (MGCP) are being proposed in an effort to simplify the implementation in the gateway and provide better scalability.

Proprietary IP Telephones
Existing PBX manufacturers are adding embedded VOP gateway functionality into their PBXs in the form of VOP line cards. This allows the PBX manufacturer to offer VOP services while still protecting the installed base of software and existing telephone sets. Over time, these PBX manufacturers will offer packet-based IP Telephone sets. To keep the cost of the telephone sets down and to take advantage of existing PBX features, manufacturers may elect to implement proprietary signaling between their IP Telephones and the PBX and perform interoperability (H.323, FRF.11, VToA) functions in the PBX. The telephone sets would still perform conversion of the speech into packets, handle echo cancellation and packet play-out functions. Signaling between the telephone set and the PBX would be based on the existing (proprietary) protocols. Interoperability between manufacturers IP Telephones would only take place at the gateway level, i.e., the ability to mix and match IP Telephone sets from different manufacturers on the same LAN would not exist.

5. Summary

The migration to data centric networks has already begun. Telecom companies reported that the volume of data traffic on public voice networks surpassed voice in 1997. Major networks are being constructed using a data infrastructure and deployment of high-speed, permanent IP connections to residences in the form of cable and xDSL modems will occur in significant volumes over the next two years. Central office equipment will migrate to hybrid data architectures to handle this new traffic. With the advent of packet voice, any device that is network enabled can be voice enabled. The IP Telephone will appear on the scene to address a wide variety of applications.

The implementation of an IP Telephone requires a great deal of sophisticated real-time software to address QOS over a data network, interoperability, and manageability. The use of new, low-cost, high performance processors combined with system on a chip integration is required to deliver this software in a cost effective solution.

About the Author

Mr. Witowsky has nearly 20 years experience in developing telecommunications products and is a founder of Telogy Networks, Inc. Mr. Witowsky is Senior Vice President of Engineering and Chief Technical Officer of Telogy where he is involved in development of embedded communications software for providing voice/fax/data over packet networks. Prior to Telogy, Mr. Witowsky was Director of Engineering at Hughes Network Systems and was involved in the development of X.25 packet switching equipment and satellite communications systems. Mr. Witowsky received his BSEE from Stevens Institute of Technology and his MSCS from Johns Hopkins University. Mr. Witowsky holds a number of patents in the field of communications and software.