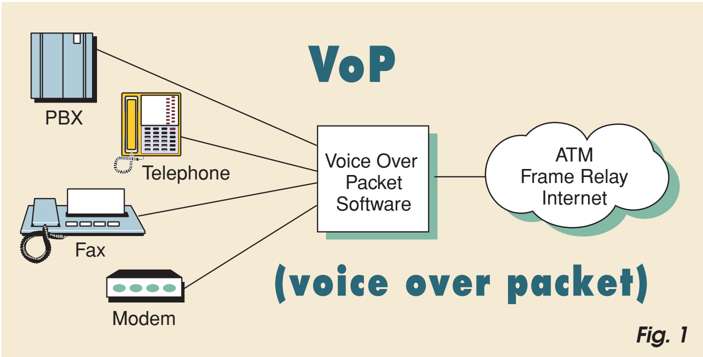

Fig. 1

Deepak Sharma

When one cannot invent, one must at least improve.

fortune cookie saying posted on blackboard at Univ. of Texas ('92)

Organizations around the world want to reduce rising communications costs. The consolidation of separate voice and data networks offers an opportunity for significant savings. Accordingly, the challenge of integrating voice and data networks is becoming a rising priority for many network managers. Organizations are pursuing solutions that utilize excess capacity on broadband networks for voice and data transmission, as well as utilize the Internet and company Intranets as alternatives to costlier mediums.

A Voice over Packet (VoP) application meets these challenges by allowing both voice and signaling information to be transported over the packet network. This technology was conceived in the early 1960s. But its real potential only recently has been understood due to an enormous rise in Communication demands. This rise in demand saw the development of two fast technology standards in the 80s in the form of Frame Relays (FR) and Asynchronous Transfer Mode (ATM). Many corporations have long been using voice over Frame Relay to save money by utilizing excess Frame Relay capacity. However, the Internet Protocol's rising dominance has shifted most attention from VoFR to VoIP.

What is VoP?

"VoP" stands for Voice over Packet. Voice is carried as digital data, often compressed, along with non-voice data over a common packet-switched infrastructure. As shown in Fig 1, the legacy telephony terminals that are addressed range from standard two wire Plain Old Telephone Service (POTS) and Fax Terminals to digital and analog PBX interfaces. Packet networks supported are ATM, Frame Relay and the Internet.

Let us take an example in which Voice over Packet software works with an existing communication network, like a cellular network, as shown in Fig. 2. The voice data in a digital cellular network is already compressed and packetized for transmission over the air by the cellular phone. Packet networks can then transmit the compressed cellular voice packet, saving a tremendous amount of bandwidth. The IWF provides the transcoding function required converting the cellular voice data to the format required by the Public Switched Telephone Network (PSTN).

Characteristics of VoP

Delay. Delay causes two problems: echo and talker overlap. Echo is caused by the signal reflections of the speaker's voice from the far end's telephone equipment back into the speaker's ear. Echo becomes a significant problem when the round trip delay becomes greater than 50 milliseconds (msec). Echo is perceived as a significant quality problem. Thus, Voice over Packet systems must address the need for echo control and implement some sort of echo cancellation.

Talker overlap (or the problem of one talker stepping on the other talker's speech) becomes significant if the one-way delay becomes greater than 250 msec. The end-to-end delay budget is therefore the major constraint and driving requirement for reducing delay through a packet network.

The following are possible delay sources in an end to end Voice over Packet call:

1. *Accumulation delay (sometimes called algorithmic delay):* This delay is caused by the need to collect a frame of voice samples to be processed by the voice coders. It is related to the type of voice coders used and varies from a single sample time (.125 microseconds) to many milliseconds. Each algorithm requires that different amounts of speech be buffered prior to the compression. This delay adds to the overall end-to-end delay (see discussion below). A network with excessive end-to-end delay, often causes people to revert to a half-duplex conversation ("How are you today? Over…") instead of the normal full-duplex phone call. (See Table 1)

2. *Processing delay:* This delay is caused by the actual process of encoding and collecting the encoded samples into a packet for transmission over the packet network. The encoding delay is a function of both the processor execution time and the type of algorithm used. Often, multiple voice coder frames will be collected in a single packet to reduce the packet network overhead. For example, three frames of G.729 code words, equaling 30 milliseconds of speech, may be collected and packed into a single packet.

3. *Network delay*: This delay is caused by the physical medium and protocols used to transmit the voice data, and by the buffers used to remove packet jitter on the receiver side. Network delay is a function of the capacity of the links in the network and the processing

Compression Scheme	Compressed Rate (Kbps)	Required CPU Resources	Resultant Voice Quality	Added Delay
G.711 PCM	64 (no compression)	Not Required	Excellent	N/A
G.723 MP-MLQ	6.4/5.3	Moderate	Good (6.4) Fair (5.3)	High
G.726 ADPCM	40/32/24	Low	Good (40) Fair (24)	Very Low
G.728 LD-CELP	16	Very High	Good	Low
G.729 CS-ACELP	8	High	Good	Low

that occurs as the packets transit the network. The jitter buffers add delay. They are used to remove the packet delay variation that each packet is subjected to as it transits the packet network. This delay can be a significant part of the overall delay since packet delay variations can be as high as 70-100 msec in some Frame Relay networks and IP networks.

Jitter. The delay problem is compounded by the need to remove jitter: a variable inter-packet timing caused by the network a packet traverses. Removing jitter requires collecting packets and holding them long enough to allow the slowest packets to arrive in time to be played in the correct sequence. This causes additional delay.

The two conflicting goals of minimizing delay and removing jitter have engendered various schemes to adapt the jitter buffer size to match the time varying requirements of network jitter removal. This adaptation has the explicit goal of minimizing the size and delay of the jitter buffer, while at the same time preventing buffer underflow caused by jitter.

The approach selected will depend on the type of network the packets are traversing. Two approaches to adapting the jitter buffer size are:

1. To measure the variation of the packet-level in the jitter buffer over a period of time, and incrementally adapt the buffer size to match the calculated jitter. This approach works best with networks that provide a consistent jitter performance over time, such as ATM networks.

2. To count the number of packets that arrive late and create a ratio of these packets to the number of packets that are successfully processed. This ratio is then used to adjust the jitter buffer to target a predetermined allowable late packet ratio. This approach works best with the networks with highly variable packet inter-arrival intervals, such as IP networks.

In addition to the techniques just described, the network must be configured and managed to provide minimal delay and jitter, enabling a consistent quality of service.

Lost packet compensation. Lost packets can be an even more severe problem. It depends on the type of packet network that is being used. Because IP networks do not guarantee service, they will usually exhibit a much higher incidence of lost voice packets than ATM networks.

In current IP networks, all voice frames are treated like data. Under peak loads and congestion, voice frames will be dropped equally with data frames. The data frames, however, are not time sensitive and dropped packets can be appropriately corrected through the process of retransmission. Lost voice packets, however, cannot be dealt with in this manner.

Some schemes used by Voice over Packet software to address the problem of lost frames are:

1. Interpolate for lost speech packets by replaying the last packet received during the interval when the lost packet was supposed to be played out. This scheme is a simple method that fills the time between non-contiguous speech frames. It works well when the incidence of lost frames is infrequent. It does not work very well if there are a number of lost packets in a row or a burst of lost packets.

2. Send redundant information at the expense of bandwidth utilization. The basic approach replicates and sends the nth packet of voice information along with the (n+1)th packet. This method has the advantage of being able to exactly correct for the lost packet. However, this approach uses more bandwidth and also creates greater delay.

3. A hybrid approach uses a much lower bandwidth voice coder to provide redundant information carried along in the (n+1)th packet. This reduces the problem of the extra bandwidth required, but still fails to solve the problem of delay.

Echo compensation. 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 pairs) 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. Also, 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 always used.

Echo is generated 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.

Technology developments & paradigms

As mentioned earlier, the technology has its roots in the early 60s. Unfortunately, back then, the large copper analog and asynchronous infrastructure caused huge amplitude and group-delay distortions and data conduction was done through "dumb" asynchronous terminals. But today we have smart and synchronous terminals as well as digital techniques that minimize the hassles of handling traffic on a node to node basis as well as increase the speed manifolds.

Frame relay and asynchronous transfer mode. Two fast packet technology standards emerged in the late 1980s: Frame Relay (FR) and Asynchronous Transfer Mode (ATM).

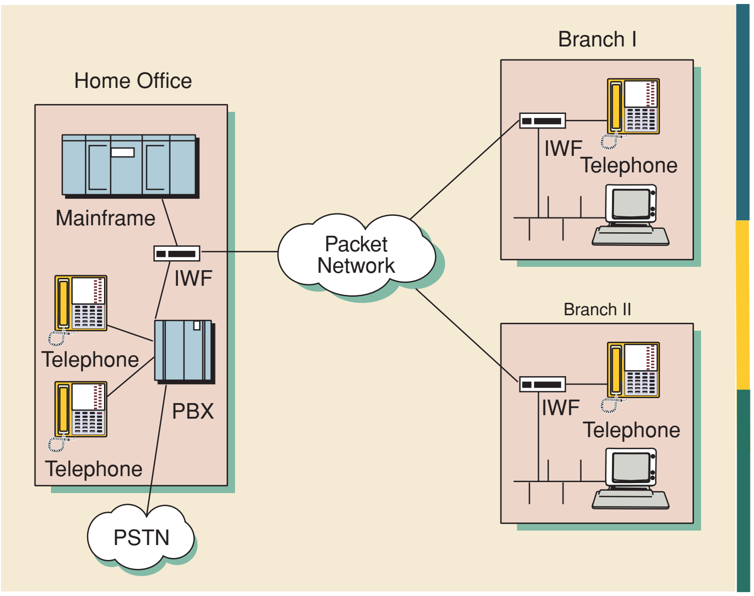

Fig. 2 Application 1: branch office

Table 2	Frame Relay	Asynchronous Transfer Mode
Packet Delay	Variable	Deterministic
Packet Jitter	Variable	Deterministic
Overhead	Low	High
Access Speed	Typical Nx64 KBPS, T1/E1	DXI, (> or =) 45 MBPS

While briefly there was confusion as to the relative positioning of these two standards, today each has a clearly defined position: FR has become the standard for low-speed access and feeder networks; ATM has become the standard for high-speed connection and backbone infrastructure.

Interoperability between these two standards has also been established. An interface known as Frame User to Network Interface (FUNI) allows a signal to enter the fast packet environment via a Frame Relay Access Device (FRAD) and be extracted by an ATM device. With this capability, the future fast packet-based telecommunications hierarchy has taken shape.

In time, fast packet technology probably will replace the existing circuit-switched infrastructure. Furthermore, in the telecommunications environment, FR and ATM exhibit attributes not seen in other network technologies: carriers around the world embrace FR and ATM. As a result, their respective growth rates are unprecedented. A brief comparison between these two technologies is shown in Table 2.

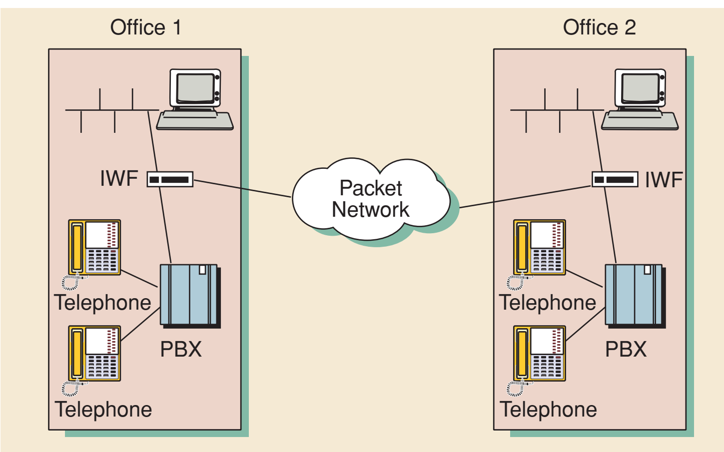

Fig. 3 Application 2: interoffice trunking

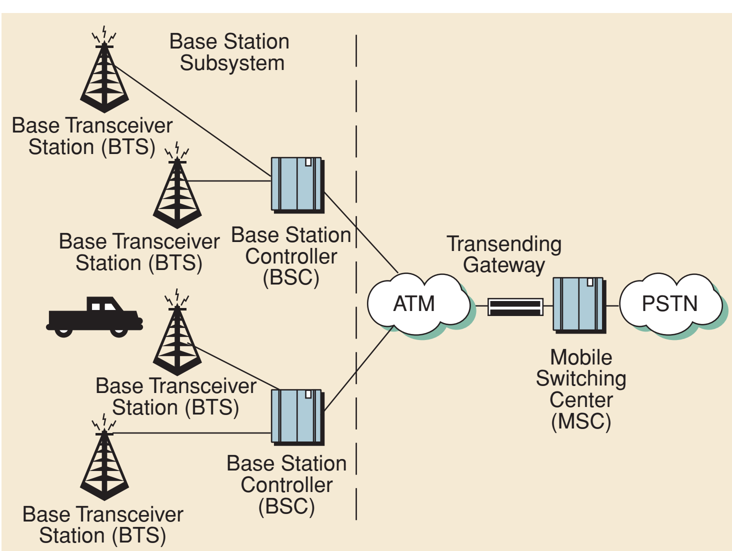

Fig. 4 Application 3: cellular networking interworking

Benefits of VoP networks

Less bandwidth consumption: Packetized voice offers much higher bandwidth efficiency than circuit-switched voice because it does not take up any bandwidth in listening mode or during pauses in a conversation. Say we were to use the same 64 Kbps Pulse Code Modulation (PCM) digital-voice encoding method in both technologies. We would see that the bandwidth consumption of the packetized voice is only a fraction of the consumption of the circuit-switched voice. For this reason, packetized voice has already been deployed in most Trans-Pacific and Trans-Atlantic Digital Circuit Multiplication Equipment (DCME).

Industrial-voice advantages: There is a number of voice products that are extremely price or cost sensitive. For example, 800 numbers are paid for by one party, 100 calling is free to the public, and inexpensive telephone directory service is key to retaining customer loyalty. Carriers could offer these types of voice products using their lower-cost fast packet infrastructure.

Low-cost PSTN options: Remember the days of pulse-dial and touch-tone services? They were offered as alternatives to consumers at different pricing levels. The same strategy can be used to move voice service from circuit switched to packet switched. A consumer can, with the touch of a button at call initiation, decide whether to call on the existing PSTN or the new packetized PSTN.

Concerns about VoP networks

Some believe that fast packet technology works well with data but not with voice. The common concerns surrounding voice over frame relay are voice quality, modem handling, network latency, congestion, switched connectivity and the commercial implications of implementing fast packet systems.

Voice quality: The earlier voice-compression systems did produce less-than-desirable voice quality. Consequently, use of these early systems was restricted mainly to secured voice applications. This legacy took years to overcome. Today, however, advanced algorithms not only have reduced the complexity of the design but they also produce voice quality equal to the standard PCM and Adaptive PCM (ADPCM) so widely deployed in the digital network.

Modem handling: The modem was invented to allow digital data to be transmitted across an analog network. It is therefore a transitional product. Ultimately, modems will be eliminated due to an ubiquitous digital infrastructure.

Packetized voice systems handle modem signals by detecting signal presence, demodulating the signal back to its original digital form, and then passing it transparently across the network. At the remote end, this data is then remodulated before presenting it to the receiving modem. A side benefit is bandwidth efficiency. Instead of a 64 Kbps PCM channel to handle a modem signal carrying 9.6 Kbps data, only 9.6 Kbps is needed.

Network latency: The concern over network latency is the result of another legacy-related misconception. Fast packet systems offer faster transmission than do older packet systems.

The old X.25 packet system uses node-to-node significance to transmit data: errors are checked every step of the way. The network, itself, can introduce a transmission delay of a few hundred milliseconds from source to destination.

Fast packet systems, on the other hand, care only about where to route the traffic, not about the integrity of the transmission. They use end-to-end significance to transmit the data. The delay from source to destination is at least two orders of magnitude less than X.25.

Congestion management: TDM/circuit-switch network design is easy to understand because it is deterministic in nature. Once bandwidth is assigned to a given application, it is protected. Such simplicity, however, is also the cause of its inefficiency.

Virtual circuit/packet switch network design, on the other hand, is statistical in nature. Bandwidth is not given until needed. This design technique is more difficult to understand and, therefore, has been less readily accepted. But, in reality, the design is quite simple and logical.

The proactive approach prevents congestion by understanding the traffic flow, properly sizing the network and constantly monitoring the bandwidth utilization. In addition, the virtual circuit/packet switch network removes the likelihood of a logjam by taking advantage of the congestion notification mechanism available in the fast packet infrastructure. Short-term bandwidth demand can be throttled back by a combination of reducing the voice encoding rate, slowing the data transfer speed and allowing certain packets to be discarded.

Switched connectivity: Packet systems offer a far more flexible way of forming switched circuits across a network than circuit systems. Each packet has a header containing the originating and destination addresses. This allows connections to be set up on a packet-by-packet basis, hence, the term virtual circuit.

Two types of virtual circuits are available: Permanent Virtual Circuit (PVC) and Switched Virtual Circuit (SVC). PVC is equivalent to today's leased line, whereas SVC is equivalent to today's switched line. SVC is now a standard; equipment vendors for switches and access devices have begun offering such a facility in their products.

Software architecture

Two major types of information must be handled in order to interface telephony equipment to a packet network: voice and signaling information (Fig. 5). Voice over Packet software interfaces to both streams of information from the telephony network and converts them to a single stream of packets transmitted to the packet network.

The software functions are divided into four general areas:

Voice packet software module: This software, typically runs on a digital signal processor (DSP), prepares voice samples for transmission over the packet network. Its components perform echo cancellation, voice compression, voice activity detection, jitter removal, clock synchronization and voice packetization.

Telephony signaling gateway software module: This software interacts with the telephony equipment translating signaling into state changes used by the Packet Protocol Module to set up connections. These state changes are on-hook, off-hook, trunk seizure and so forth.

Packet protocol module: This module processes signaling information and converts it from the telephony signaling protocols to the specific packet signaling protocol used to set up connections over the packet network (e.g. Q.933 and Voice over FR signaling). It also adds protocol headers to both voice and signaling packets before they are transmitted into the packet network.

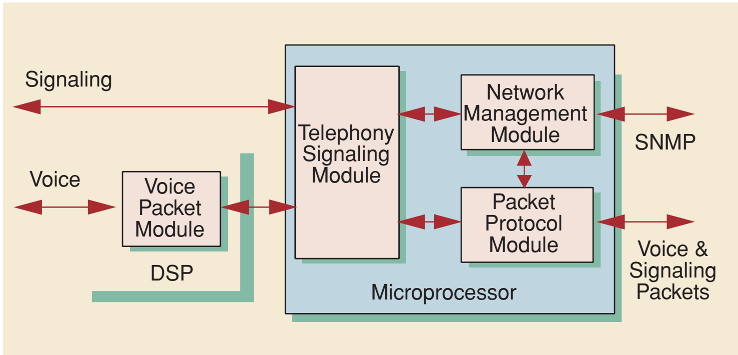

Fig. 5 Voice over Packet software architecture

Network management module: This module provides the Voice management interface to configure and maintain the other modules of the voice over packet system.

The software is partitioned to provide a well-defined interface to the DSP software usable for multiple voice packet protocols and applications. The DSP processes voice data and passes voice packets to the microprocessor with generic voice headers.

The microprocessor is responsible for moving voice packets and adapting the generic voice headers to the specific Voice Packet Protocol called for by the application, such as Real-Time Protocol (RTP), Voice over Frame Relay (VoFR), and Voice Telephony over ATM (VTOA). The microprocessor also processes signaling information and converts it from supported telephony signaling protocols to the packet network signaling protocol (e.g. H.323 (IP), Frame Relay, or ATM signaling). This partitioning provides a clean interface between the generic voice processing functions, such as compression, echo cancellation, and voice activity detection.

Conclusion

All the major communication companies are pouring money into the research of carrying voice in a digital compressed form along with the non-voice data over a common packet switched network. Much work still needs to be done. Better technology is required for, among other things, minimizing delay and reducing transmission errors.

Voice quality alone cannot guarantee the success or failure of voice over packet networks. All the same, speech quality remains one of the prime hurdles affecting such networks.

Acknowledgements

Thanks to my teachers who helped me prepare this paper, especially my project guide, Prof. S.L. Haridas and Prof. M.A.Gaikwad. Thanks to the college authority for allowing me to access the college Internet for my research. Finally, thanks to the librarian who arranged the encyclopedias on *IEEE Spectrum* magazine for me .

Read more about it

• *IEEE Spectrum*, February 2000
• *IEEE Spectrum*, March 1998.
• IEC ONLINE EDUCATION (www.iec.org/online/tutorials).
• www.protocols.com/papers/voip 2.htm
• www.telogy.com/our_products/golden_gateway/VOPwhite.html

About the author

Deepak Sharma received his Bachelors in Electronics Engineering from BD College of Engineering, Nagpur University in India.