

VOIP IN THE BROADCAST STUDIO

Michael Dosch and Steve Church

Axia Audio/Telos Systems, Cleveland, OH

Without much doubt, VoIP (Voice over Internet Protocol telephony) is coming to broadcast facilities. We explain why this is so, the benefits and downsides, and how the systems will work and integrate with studio audio equipment.

There are three distinct possibilities for applying VoIP in a broadcast studio or audio production facility:

- Using an IP-based PBX for general phone service
- Using VoIP to connect to the telco network
- Using an IP-based studio telephone system for on-air calls

First, we'll survey what is happening in the world of telephony at large, and then we'll move on to what it means for modern broadcast studio design

IP PBXS

As anyone who has noticed the proliferation of Cisco-branded and other IP phones on business desktops could tell you, VoIP PBXs are rapidly replacing the old-style TDM (Time Domain Multiplex) proprietary ones, especially in large organizations.

Reportedly 80% of all new PBX lines installed worldwide in 2008 were VoIP.

VOIP TELCO SERVICE

In a petition [1] that will probably come to be regarded as historic, AT&T has asked the FCC to order the shutdown of the PSTN (Public Switched Telephone Network) – the ubiquitous telephone system that provides POTS, T1, and ISDN switched voice service.

In a public response to the US Federal Communications Commission's request for comments regarding its forthcoming National Broadband Plan, AT&T acknowledged not the future obsolescence, but the *current* obsolescence of the PSTN telephone system, the one-time marvel of technology that defined its predecessor, the Bell System, in the 20th century.

AT&T wrote:

Due to technological advances, changes in consumer preference, and market forces, the question is when, not if, POTS service and the PSTN over which it is provided will become obsolete.

In the paper, AT&T credited the success of Skype and Vonage for having driven up subscriptions to VoIP service, which it now believes to be 18 million subscribers -- a number that it believes could triple in two years' time. But the impressive statistics cited by AT&T do not end there. Today, fewer than 20% of Americans rely exclusively on POTS for voice service. Approximately 25% of households have abandoned POTS altogether, and another 700,000 lines are being cut every month. From 2000 to 2008, the number of residential switched access lines has fallen by almost half, from 139 million to 75 million. Non-primary residential lines have fallen by 62% over the same period. Total interstate and intrastate switched access minutes have fallen by 42% from 2000 through 2008. Indeed, perhaps the clearest sign of the transformation away from POTS and towards a broadband future is that there are probably now more broadband connections than telephone lines in the United States. And the customers who keep POTS are using it less. Wireless phones, e-mail, instant messaging, blogs, and social networking sites have greatly reduced the need for legacy voice services, even for customers who retain POTS service. Between 2000 and 2008, switched access minutes per line declined by 13.2%.

AT&T says that the funds they would have to spend to maintain the PSTN should be spent instead to build out a ubiquitous IP broadband system. Presumably, this would include a significant upgrade to their mobile network that would strengthen their existing infrastructure to better compete with carriers that don't have expenses associated with "legacy" obligations.

The story is likely to be similar outside of the USA. While researching the VoIP chapter for the book Skip Pizzi and Steve co-authored [2] we heard from one European telco that they had stopped putting money into their TDM network altogether and were beginning a full transition to IP. Subscribers who insist on having an analog phone jack will get an IP-to-POTS adapter box.

As a side note, last time we checked, the equipment-making stepchild of the formerly vertically-integrated AT&T, Alcatel/Lucent, had no TDM-based central office products on their website. Instead, they are promoting their IP-based IMS systems.

You might be thinking, "Does it really make sense to rip out the PSTN that has served us reliably for decades?" Seems so. With content providers and users

hungry for bandwidth, IP broadband is the fastest and cheapest way to provide it.

AT&T can get 100Mbps out of a copper line that's not too long and in good condition. Feeding those with IP is much cheaper and simpler than with discrete 64kbps speech channels. And you get much more flexibility. Fast web access, video, and hi-fi voice are the already demonstrated applications, but more are sure to appear. The network also benefits from the efficient use of bandwidth that statistically-multiplexed IP routing offers compared to fixed-size circuit-switched channels.

At the dawn of the Internet, early adopters were running 1.2kbps modems over a network built for lo-fi analog voice. These days voice is very often run over an IP network as just another Internet application ("JAIA" in the emerging jargon).

AOIP IN BROADCAST FACILITIES

Meanwhile, Audio over IP (AoIP) is driving a revolution in audio studio design, replacing traditional purpose-built mixers, routers and switchers with an architecture that's more computer-friendly, more scalable, faster to install and future-proof.

VOIP APPLICATIONS IN RADIO/TV

With VoIP PBXs becoming widespread, VoIP telco service on the horizon, and AoIP quickly taking hold for building studio infrastructure, IP-based on-air telephone systems can't be far behind.

VoIP phone systems and AoIP studio networks can be tightly interconnected, creating numerous benefits with regard to ease of installation and support of desirable features. Unnecessary analog-to-digital and 2-to-4 wire conversions are eliminated, allowing calls to pass cleanly over the studio network for better-sounding calls. Much development is taking place in the VoIP world, and some of it has strong bearing on the technology's application to broadcasting and other audio production facilities. As the technology matures, so should broadcasters' awareness of it, such that its advantages can be put to proper use in radio and television production systems.

Advantages of VoIP for Broadcast Facilities

Consider the following commonly encountered telephone-related problems at broadcast/audio facilities.

- 1) Transferring calls between a facility's office PBX and its studio telephone-interface lines is cumbersome.

- 2) It's cost-effective to have a single, multi-line digital connection to your telco. But it's not easy to subdivide incoming lines between the facility's PBX and its studio telephone interfaces.
- 3) Multi-studio facilities can benefit from sharing a single line pool among the studios. But this is often difficult and/or expensive to achieve.
- 4) Calls from mobile phones have audio quality bordering on unacceptable for on-air use.

VoIP has the potential to solve these problems. In particular, there is the opportunity for wideband, near hi-fi connections from both mobile and fixed-line callers.

Proprietary PBXs

The first trouble in the list arises from the proprietary digital formats used by today's typical PBXs. With no standardized protocol available for studio equipment and PBXs to communicate with each other, analog ports are often the only way to interconnect. The limitations of the primitive signaling possible over these connections mean that even basic information such as is-the-line-on-hold? cannot be conveyed.

As well, calls transferred from the PBX to the studio are forced to go through an unnecessary low-grade analog-to-digital and digital-to-analog conversion, which add noise and distortion.

Proprietary PBX formats also cause the second and third troubles listed above, preventing your broadcast facility from taking advantage of the cost and quality benefits of direct, multi-line digital telco services (such as T1 or ISDN-PRI lines).

In the IP world, both audio formats and a control protocol have been standardized so that compatible equipment can interconnect and interoperate. The audio format is usually referred to as *Real Time Protocol* (RTP) and the control protocol is *Session Initiation Protocol* (SIP)

SESSION INITIATION PROTOCOL

SIP is an IETF standard¹ used to establish calls over IP connections. It enables familiar, telephone-like operations such as dialing a number, causing a phone to ring, sending ringback tones or a presenting busy signal. It also enables next-generation, "smart" capabilities such as finding people and directing calls to them at any location, Instant Messaging, and

¹ SIP was first specified by the IETF's RFC 2543 (March 1999), improved and expanded in RFC 3261 (June 2002).

relaying so-called “presence” data (e.g., near the phone or not, do-not-disturb, etc.).

SIP began as a simple message protocol for setting up connections, but now the term has grown to be an umbrella for the family of protocols and tools that have been developed by the IETF to enable VoIP telephony and related services.

SIP’s standardized approach makes it simple and routine to hand calls back and forth between a station’s office and on-air systems, and allows smooth, interoperable communication among different vendors’ equipment. Further, the use of IP transport throughout avoids unnecessary A/D-D/A conversions and lets telephone audio pass in pure digital form throughout the signal path.

For telco connections, a gateway can be used to interface POTS, T1, or ISDN to the VoIP system. A PBX can serve as a gateway. Or both the station’s office PBX and the studio on-air lines can use SIP to interface to a gateway. The channels (“phone lines”) assigned by telco to the station can be divided any way that the station desires.

SIP Trunking

SIP Trunking is roughly equivalent to a T1 or primary ISDN line. A single IP connection supports multiple number presences and audio channels. This will usually be how a studio on-air phone system will interface to a gateway or IP PBX. It will also be how VoIP telco service connections will be made.

How SIP Works

Like HTTP (Hypertext Transfer Protocol), SIP is human readable and request-response structured. SIP also shares some of HTTP’s status codes, including the well known “404 not found.”

The following is an example of a SIP message:

```
INVITE sip: mike@there.com SIP/2.0
Via: SIP/2.0/UDP 4.3.2.1:5060
To: Mike Dosch<sip:mike@there.com>
From: Steve Church <sip:steve@here.com>
Call-ID: 4678995554545@4.3.2.1
CSeq: 1 INVITE
Contact: <sip:steve@4.3.2.1>
Content-Length: 126
```

That message would indicate to Mike’s SIP client that Steve’s client wants to connect.

SIP is only one of several protocols used in VoIP. Signaling duties for a communication session are handled by SIP, and it serves as a carrier for the *Session Description Protocol* (SDP),² which describes the media content of the session, such as the codec used, the bit rate, and the like.

The following list summarizes the capabilities of SIP:

- SIP determines the location and availability of the called location. It supports address resolution, name mapping, and call redirection. If a call cannot be completed because the target endpoint is unavailable, SIP returns a message indicating this and why.
- If the call can be completed, SIP establishes a session between the originating and called endpoints.
- SIP determines the media capabilities of the endpoints, including which codecs are supported, and negotiates with the called endpoint to use the most appropriate codec for the call.
- SIP handles the transfer and termination of calls. For a call transfer, SIP establishes a new session between the transferee and a new endpoint (specified by the transferring party), and terminates the original session.

SIP System Components

SIP uses a modular design, as does almost all IP-based networking. Systems can be built from any of the following components:

- **SIP Client:** Also known as *User Agents* or *Endpoints*, these are implemented either in a telephone hardware set or as a “softphone,” which is a telephone application that runs on a PC.
- **Registrar server:** A type of server that processes requests from SIP clients for registration of their current location.
- **Redirect server:** A type of server that presents SIP clients with information about the next networking routing segment(s) that a message should take. This permits the SIP client to contact the next server or SIP client directly.
- **Proxy server:** A type of server that receives requests from a SIP client and forwards them to the next SIP server in the network. Such servers can provide authentication, authorization, network access control, routing, reliable request retransmission and security functions.

² Originally defined in IETF’s RFC 2327 (April 1998), revised in RFC 4566 (July 2006).

- **Gateway:** This device provides physical, electrical, signaling and audio interfaces between the IP domain and the switched-circuit telco domain.

SIP clients can connect to each other directly, but the SIP servers above provide some additional, desirable functions:

- They can register SIP client devices
- They can look up the address of the far endpoint
- They register individual human users for access to VoIP services
- They can provide user mobility across networks and devices
- They can support multipoint conferencing, presence information and call progress details.
- They can request QoS data from other network elements (e.g., IP routers)
- When needed, they can provide authentication, authorization and accounting functions

The individual servers listed above can run independently and be physically separated, but they are often combined into an application that runs on a single machine. Some IP-based PBXs also include a Gateway described above, thereby providing a one-box solution for a small-office installation.

One example of a SIP Server in use by broadcasters is the Telos *Z/IP Server*. The server is typically provided as an Internet service, but it can also be locally installed within a facility. Besides basic SIP functions of registration and address look-up, the *Z/IP Server* also offers the following:

- It provides geolocation services by associating IP numbers with physical locations, and displays a routing map.
- It holds a user database of names, and allows display and dialing by simple text name; it performs Domain Name System (DNS)/IP look-up upon a dialing request from an endpoint codec. It also allows users to create group lists, which can be displayed on endpoint codecs.
- Upon request, it keeps a record of network performance in order to assist in troubleshooting problems caused by Quality of Service (QoS) impairments.

Note that many products supporting SIP for its standards-based interconnection capability do not have internal architectures that fully adhere to the SIP standard, so these SIP server components might not be included in product specs, or may be listed under other

proprietary names. Cisco and Microsoft make popular VoIP products that follow this approach.

SIP Addressing

SIP addresses, or SIP URIs (Uniform Resource Identifiers), take the following form:

```
sip:user@host
```

User can be a text name or a telephone number, and host is a domain name or IP address. Generally, SIP address resolution uses this URI to arrive at a username, at an IP address. No information about the physical location or IP address of the receiver is needed (as with e-mail). Thus SIP can automatically implement mobility and portability.

Some examples of valid SIP addresses are presented below.

The usual form is an email address prefixed by sip:, as follows:

```
sip:joesmith@company.com
```

This is how to call a PBX telephone at an enterprise (e.g., extension 123 at Telos Systems):

```
sip:123@telos-systems.com
```

If the caller doesn't know a name or extension, the receptionist can be contacted:

```
sip:receptionist@telos-systems.com
```

An internal machine-to-machine message, such as from an on-air phone system to a PBX or gateway to initiate a PSTN³ call, takes this form:

```
sip:12162417225@168.123.23.1
```

Here an IP number is provided to identify the physical device that is targeted. This avoids the use of DNS and thereby saves time. Moreover, computers used as telephone servers may not always have a DNS name associated with them.

SIP allows the use of the +, -, and . characters as separators, to assist human readability. (These characters are removed prior to processing.) The following is a valid SIP input address:

```
sip:+1-216-241-7225@telos-systems.com
```

Thus SIP bridges the telephone and Internet worlds. Both web-style and PSTN telephone number addresses are usable, and clients on either network can reach clients on the other.

³ *Public Switched Telephone Network*, affectionately known as "POTS," for "Plain Old Telephone Service."

The SIP address resolution process usually involves multiple steps and message hops. For example a single name resolution may involve a DNS server, a SIP proxy server and a SIP redirect server.

Note that some servers associated with SIP systems can accept unformatted text names, but this is not part of the standard.

Importantly, the URIs used by SIP are not URLs (Universal Resource Locators). Remember that URIs are independent of physical location. A *Request-URI* is used to indicate the destination name for a SIP Request (INVITE, REGISTER, etc.). URLs then describe the location of a resource available on the Internet. For example, <http://www.telos-systems.com> is the URL for a Web home page. It is resolved by DNS to an actual IP address.

PSTN telephone numbers can be referred to as *E.164* numbers, which refers to an ITU-T standard⁴ of that name describing the format of telephone numbers around the world. A part of the DNS is *ENUM* (**E**.164 **N**umber **M**apping), an Internet service that looks up the URI associated with an E.164-formatted telephone number. SIP uses ENUM to locate the VoIP system associated with a telephone number that accepts incoming calls.

SIP in Practice

As illustrated above, SIP is a simple, text-based protocol. It establishes communication between various components of a network using requests and responses, and ultimately establishes a connection between two or more endpoints, as shown in Figure 1.

In actual practice, however, SIP servers of various implementations are involved. When a call is initiated, a SIP request is sent to a SIP proxy or redirect server, which includes the addresses of both the caller and the called party. Alternatively, users can register their assigned SIP addresses with a registrar server, which provides the address when a location server requests it.

Occasionally a SIP user may move between end systems. The location of the user can be dynamically registered with the SIP server. Because the end user can be logged in at more than one station, and because the location server can sometimes have inaccurate information, it might return more than one address for the end user. If the request is coming through a SIP proxy server, the proxy server tries each of the returned addresses until it locates the end user. If the request is coming through a SIP redirect server, the

redirect server forwards all the addresses to the caller in the Contact header field of the invitation response.

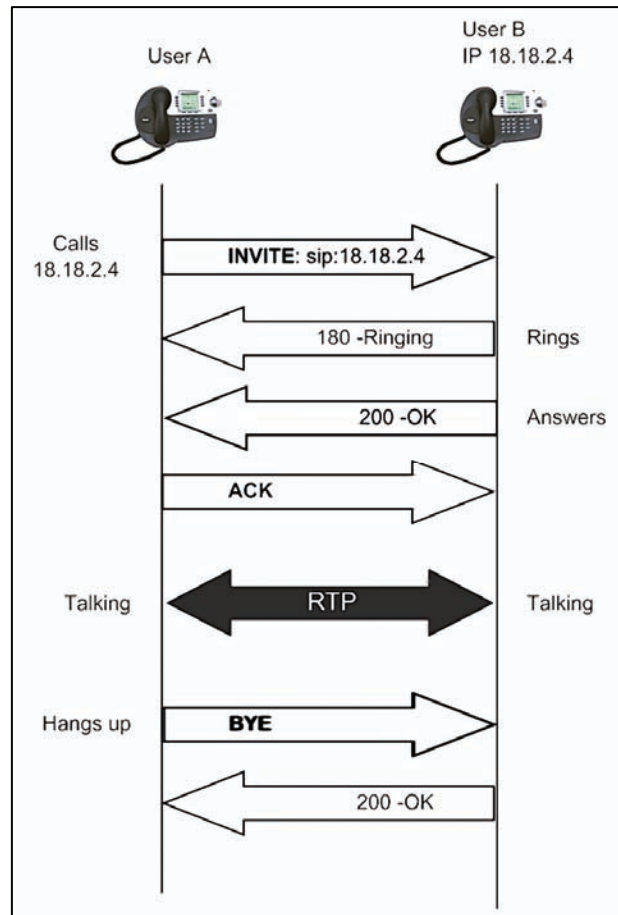


Figure 1: The most basic SIP functionality, in which one IP phone calls another IP phone directly.

If a caller is working through a proxy server, the INVITE request is sent to the proxy server, and the proxy server determines the path, then forwards the request to the called party.

Because the end targets are often phones connected to the PSTN, gateways typically will be involved in real-world systems. These translate SIP signaling to the PSTN's requirements for the last mile: loop current, tone- and ring-generation/detection, set-up messages for ISDN, etc. An example of a call set-up involving such interconnection is shown in Figure 2.

SIP messages may be carried by TCP⁵

⁴ ITU-T Recommendation E.164 (02/2005): *The International Public Telecommunication Numbering Plan*

⁵ *Transmission Control Protocol*, one of the Transport Layer ("Layer 4") protocols used in the Internet networking model. TCP includes substantial error-correction capabilities, at the expense of potentially increased latency (i.e., throughput delay).

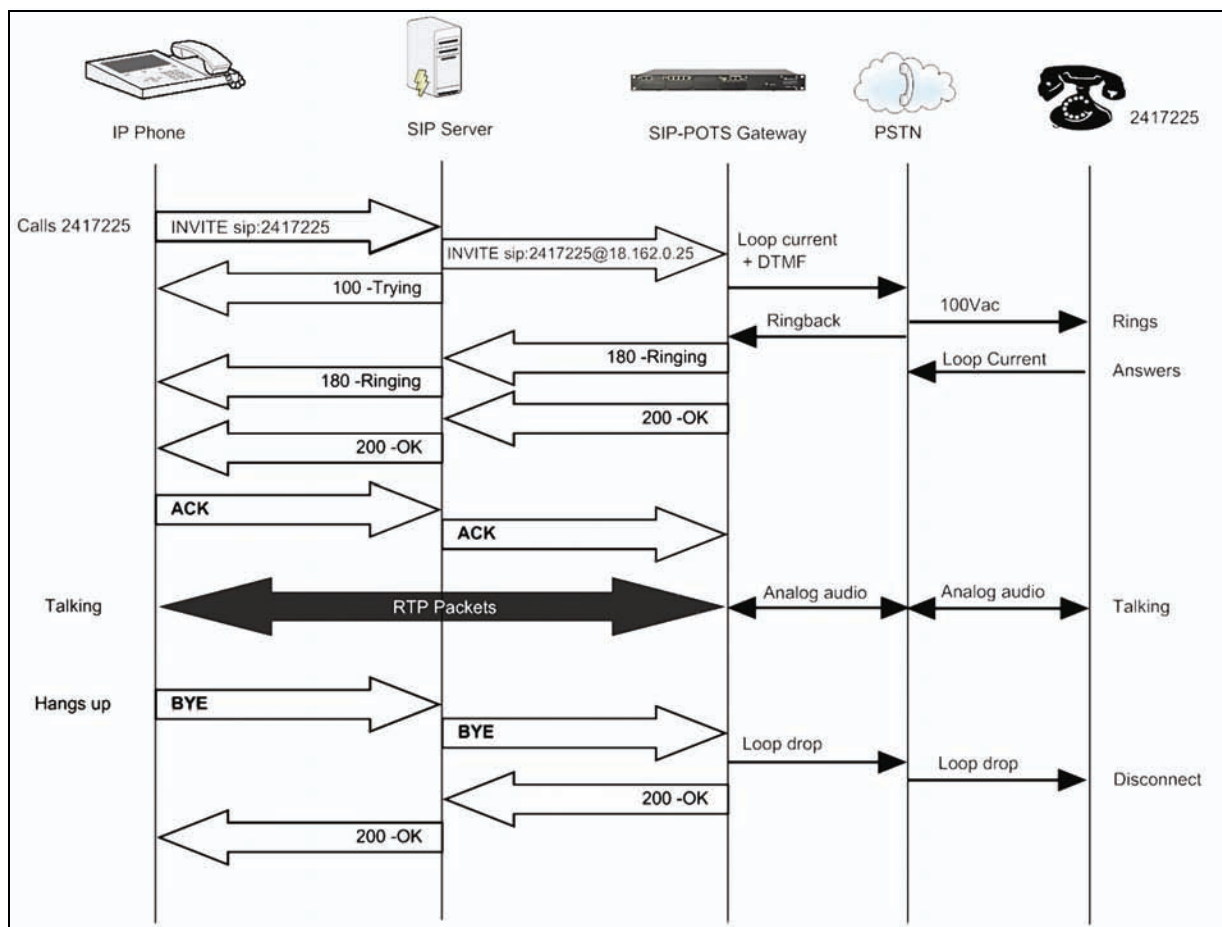


Figure 2: A SIP call set-up from an IP phone to the PSTN, using a SIP server and Gateway to “POTS” (PSTN) phone lines

or UDP.⁶ Because SIP has its own built-in reliability mechanisms, it doesn't need TCP's reliability services. Most SIP devices such as phones and PC clients therefore use UDP for transmission of SIP messages. PBXs on LANs almost always use UDP because LANs don't drop packets, and there is no need to incur the additional overhead of TCP. *Transport Layer Security (TLS) protocol*⁷ is sometimes used to encrypt SIP messages. TLS runs on top of UDP (or any Transport Layer protocol).

An additional process not shown in Figure 1 or 2 is the media negotiation that is part of the INVITE/200 OK/ACK sequence. This is how endpoints decide which audio codec to use. SDP defines how codecs are offered and accepted on IP calls. Usually, the caller

sends an SDP message along with its INVITE, listing the codecs it is prepared to use. The far end chooses one of them and tells the caller which it prefers in the 200 OK response. Alternatively, the caller can let the far end propose a codec by not sending an SDP message in its INVITE. It is possible that the two endpoints have no codec in common and the connection is unable to proceed, but systems are designed so that this rarely happens. For example, almost all phones, gateways and SIP telco services support the ITU G.711 codec, so the two endpoints should usually find common ground. Within a PBX system, designers usually choose one codec as a standard for the system and stick with it for all connections.

SIP Today and Tomorrow

Today's PBXs generally don't use SIP as it was intended by its developers, in that they do not employ SIP servers at their core. Instead they use their own rough equivalents, which have been designed independently.

This is likely due to SIP developers' desire to have the protocol support rich media, mobility, portability,

⁶ *User Datagram Protocol*, another of the Transport Layer protocols used in the Internet networking model. UDP assumes error correction will be handled by the Application Layer above it, and thereby delivers lower latency than TCP.

⁷ TLS is the successor to the Secure Sockets Layer (SSL), and is now at V1.2, as specified in IETF's RFC 5246. It is also used with HTTP to produce the HTTPS protocol for secure Web transactions.

sophisticated endpoints, and the like, while ignoring more mundane and practical considerations. For example, consumer PC-to-PC VoIP products must solve the problem of firewall and NAT⁸ traversal, which has been addressed quite slowly within the SIP working groups. Meanwhile, VoIP implementers like Skype dealt with these issues quickly and effectively. In addition, commercial vendors typically want to implement features that differentiate their products, and often prefer to do so unilaterally rather than wait for approval of a standards body.

Therefore most of today's VoIP-supporting products⁹ use proprietary protocols within the boundaries of their systems, while implementing SIP at the edges to connect with other vendors' products – and eventually to the telco network.

To date, therefore, SIP's greatest value is its ability to serve as an interoperability layer that allows various proprietary systems to work together. Studio telephone interfaces can talk with PBXs for the first time, PBXs can talk to one other, and eventually they will all be able to talk to telco networks.

Asterisk

Note that SIP is not the only VoIP interconnection method in use today. The **Inter-Asterisk eXchange (IAX)** protocol is an alternative to SIP for interconnections between both VoIP servers and for client-server communication.

IAX2, as the current version is named, uses a single UDP data stream (usually on port 4569) to communicate between endpoints, both for signaling and data. The voice traffic is transmitted in-band, in contrast to SIP, which uses an out-of-band RTP stream for audio. IAX2 supports multiplexing channels over a single link. When trunking, data from multiple calls are merged into a single set of packets, meaning that one IP datagram can deliver control and audio for more than one call, reducing the effective IP overhead without creating additional latency.

As IAX's name indicates, it was invented by the Asterisk¹⁰ consortium as a way to trunk calls between one Asterisk server and another. It has since moved

⁸ *Network Address Translator*, typically used today to allow more than one computer to share the IP address provided to a facility (home or small office, typically) by an ISP.

⁹ This includes Cisco, Microsoft, Telos Systems and almost all PBXs vendors.

¹⁰ Asterisk was one of the first software-based PBX implementations (1999), of which there are now many. Its "hybrid" open-source/proprietary model has attracted a large number of developers to the system.

beyond the Asterisk domain alone, and is now supported in a variety of softswitches and by a few VoIP carriers. Its main advantages are its bandwidth efficiency and simpler firewall configuration, since all traffic flows through a single port.

VOIP CODECS

VoIP can use a variety of codecs, and the codec used can be chosen based on the type of transport network and requirements of the application. Table 1 below shows commonly used VoIP codecs.

The packet sizes given in the table are default values, which some equipment will let the user change on some codecs. For example, G.711 is often default-set to 10ms in order to reduce latency, but this produces trade-offs. The smaller packet size that results generates more IP header overhead and thus lowers overall bandwidth efficiency. Smaller packets also consume more processing power in the equipment.

The bit rates given in Table 1 for the MPEG codecs are target rates. Useful bit rates range from 32-96kbps for AAC-LD¹¹ and 24-96kbps for AAC-ELD.¹²

VOIP PACKETIZATION

*Real-time Transport Protocol (RTP)*¹³ is a streaming-media packetization standard used in both VoIP and AoIP networks. RTP can run over either TCP or UDP, but for the same reasons noted above with SIP, VoIP systems use UDP for audio-payload IP packets (with RTP packetization to optimize the transport for real-time streaming). Again, this RTP-over-UDP avoids the increased delay and requirement for long receive buffers that RTP-over-TCP's packet-loss recovery schemes would require.

Within LANs, where there is no packet loss, UDP's lack of inherent packet recovery is not a problem. When using UDP on Wide Area Networks (WANs), however, dropped RTP-payload packets may occur and must be addressed by the audio codec, which must employ error concealment to reduce audibility of lost audio samples. This is particularly a requirement for

¹¹ The low-delay variant of MPEG Advanced Audio Coding.

¹² The extremely low-delay variant of MPEG Advanced Audio Coding.

¹³ Specified in IETF's RFC 3550 (July 2003), RTP is used for transfer of multimedia data over real-time streams, while its associated *Real-time Transport Control Protocol (RTCP)* periodically sends control information and QOS parameters.

Table 1: Codecs used by VoIP systems

Codec	Audio B/W*	Bit rate	Packet size	Bit rate after packetization	Notes
G.711 u-law	3.4kHz	64kbps	20ms	88kbps	US PSTN standard
G.711 A-law	3.4kHz	64kbps	20ms	88kbps	European PSTN standard
G.729a/b	3.4kHz	8kbps	20ms	32kbps	Common lo-fi VoIP codec
G.723.1	3.4kHz	5.3 or 6.3kbps	30ms	22.3kbps	Very low rate codec
G.726	3.4kHz	16-32kbps	20ms	40-56kbps	Better quality than G.729
G.722	7kHz	48/56/64kbps	20ms	88kbps (at 64kbps)	Wideband. Simple low-delay ADPCM codec. Now in Cisco phones.
G.722.1 Annex C	14kHz	24/32/48kbps	20ms	40/48/64kbps	Wideband. Also called Siren14. Invented by Polycom, and used in its video conferencing systems.
G.722.2/ AMR-WB	7kHz	6.6-23.85kbps	Variable	Unknown	Wideband. ITU mobile phone standard.
G711.1	7kHz	64/80/96kpbs	Variable	Unknown	Wideband extension to G.711
G.729.1	4kHz/7k Hz	8-32kbps	Variable	Unknown	Newer, scalable version of G.729. At higher rates, becomes a wideband codec.
iLBC	3.4kHz	15.2 or 13.33kbps	20 or 30ms	Unknown	Proprietary codec invented by Global IP Sound. Available in some Cisco phones.
RTAudio	3.4/ 7kHz	8.8/18kbps	20ms	39.6/58kbps (includes FEC)	Microsoft proprietary codec, used in the Office Communications product family. Has both narrow and wideband modes.
MPEG AAC-LD	20kHz	48-64kbps	10ms	96kbps (at 64kbps)	Full-fidelity. Used in some IP video conferencing products.
MPEG AAC-ELD	20kHz	32-64kbps	20ms	48kbps (at 32kbps)/ 96kbps (at 64kbps)	Newest codec in the AAC family. Full-fidelity for music and voice at low rates. Used in broadcast codecs.

* - "Audio B/W" values are actually the high-frequency cutoffs of the codecs' audio frequency response.

wireless and public Internet applications, where packet loss is a frequent occurrence.

Another solution involves the use of guaranteed QoS on any WAN IP links. This is possible on private or "virtual" private networks (VPNs) such as those that link corporate headquarters to branch offices. It is now becoming possible to order IP telephone service from telco providers with QoS guarantees.

RTP and Packet Size

In its usual form, the RTP header occupies 12 bytes. When added to the UDP (8 bytes) and IP (20 bytes) headers, as shown in Figure 3 below, a total header length of 40 bytes results.

The VoIP audio codec output is broken into segments and put into the IP packets following the headers. Some codecs are frame-based and thus have an inherent packet-ready format. For example, G.729 has

a 10ms frame, which could be placed one-to-one inside IP packets. But usually two frames are grouped together into one IP packet to improve efficiency. The MPEG codecs used by VoIP have longer frame lengths and are usually packetized one-to-one into IP packets.

Codecs such as the G.711 companded-PCM and G.722 ADPCM work on a sample-by-sample basis and have no inherent frames, so they may be packetized at any desired boundary. A 20ms packet size is often chosen as a compromise between delay and efficiency, but sometimes 10ms or 30ms is used when either lower delay or higher efficiency is preferred, respectively.

For example, using G.711, there are 80 bytes of data produced for each 10ms of audio. A 40-byte header on an 80-byte payload is possible, but the header-to-payload ratio that results is not very bandwidth-efficient. This is why much of the VoIP world has settled on 20ms packets. That results in a 40-byte header and 160-byte audio payload, which presents the reasonable compromise illustrated in Figure 3.

On LANs, bandwidth is plentiful, so efficiency is not a big concern. Thus studio-grade AoIP systems may even use configurations in which the header is larger than the audio payload. The primary concern in such AoIP systems is very low delay – much lower than the target for VoIP systems. In VoIP systems that run only over LANs, implementers can similarly decide to allow low delay take priority over efficiency, and therefore operate with smaller packets.

But for VoIP systems that run over WANs efficiency of much greater concern, for multiple reasons. First, bandwidth is expensive on private networks. Second, on the public Internet, a lower bit rate increases the likelihood of unbroken conversation on the VoIP call. To assist in this, header compression is sometimes used. An increasingly deployed method (especially for wireless VoIP) is the *Robust Header Compression* (ROHC) specification.¹⁴

THE PROMISE OF HI-FI PHONES

As both home and mobile networks transition to IP, we can expect the fidelity of telephony to improve. SIP allows codec selection on a call-by-call basis and IP is not limited to a particular bandwidth, ie the 64kbps bitrate standard throughout the PSTN or the 9.6 or 14.4kbps rate common to mobile phones.

¹⁴ There are currently two ROHC profiles defined for the compression of IP/UDP/RTP traffic – the original definition in RFC 3095, and the more recently published RFC 5225.

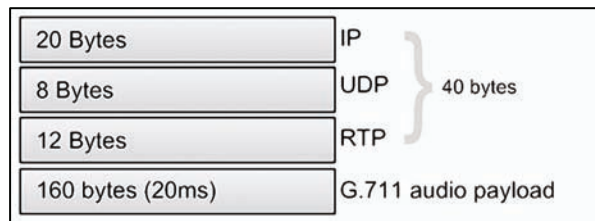


Figure 3: IP/UDP/RTP header plus 20ms of G.711 audio payload per packet is an efficiency-vs.-delay tradeoff. Each codec type may make a different tradeoff.

Mobile phone audio quality may particularly improve if the emerging AMR-WB codec or some other wideband codec gains traction. As with all things IP, this could happen on an individual phone-by-phone basis as each user decides to upgrade, not only in the case that a standards body or carrier decrees it.

Some broadcasters equate VoIP with poor audio quality, and have therefore avoided it. Most of this low-fidelity reputation is due to network QoS problems, but much was the result of the codecs that were chosen. The widely used G.729 has a bit rate of only 8kbps, and therefore cannot provide high-quality audio.

More recently, however, VoIP PBXs have settled generally on using at least G.711 for internal calls, and many are now moving up to G.722 “wideband” codecs or better. Telcos providing “business-class” IP services typically offer G.711 as their baseline codec, rather than the earlier, low-grade G.729.

Users of popular VoIP softphones such as Skype also have noticed that the service’s audio quality is better than typical phone calls, at least in terms of frequency response.

Calls from mobile phones that are eventually upgraded to a wideband codec will suffer a downgrade in fidelity when they are carried over the PSTN, but would retain their wideband quality when passed to studio systems via IP.

So in contrast to these earlier notions, VoIP is likely to be ultimately associated instead with *improved* telephone audio quality.

IP IN THE PBX

An IP-PBX system typically includes gateways to/from the IP network and PSTN or ISDN telco connections, call management software, and IP phones, as shown below. IP phones may look like traditional telephone sets, but can also be “softphones”, which are software applications running on standard PCs.

Call management software runs either on a PC or on dedicated hardware. Application servers provide any needed additional functions, such as voicemail. In some systems, these are integrated into the call management software, or can otherwise run on the same machine.

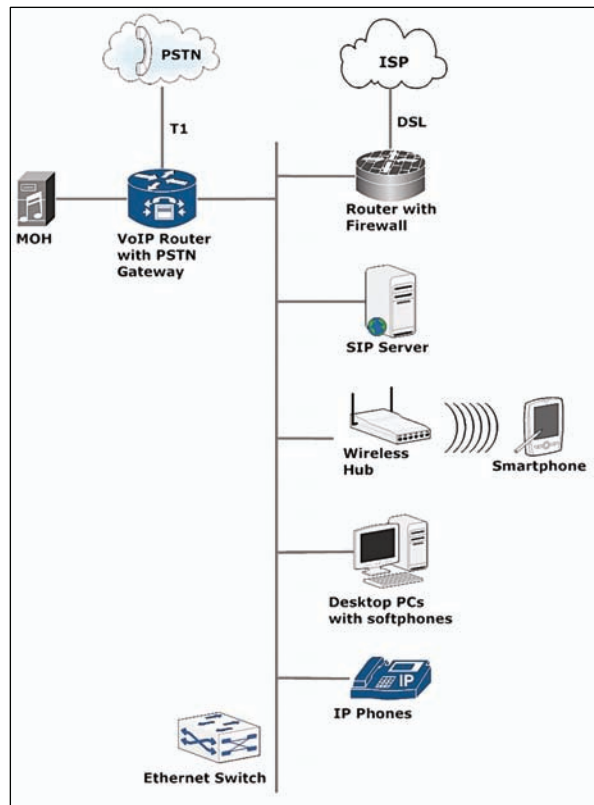


Figure 4: A SIP-based VoIP system. The VoIP router could be any of a number of available products. Calls could also be passed to an IP network directly over the firewalled DSL.

For systems where connection to telco is via SIP, no gateways are needed – the local IP network connects directly to the telco’s IP network, although a firewall may be required.

System configuration and management is performed via a Web browser pointed at one or more system elements. Most vendors use proprietary communications protocols between their call-management application and their telephone sets, ostensibly to support the features on the phones, such as displays and soft keys. Most systems also support a basic variant of standard SIP, which allows third-party SIP phones and other endpoints to be attached to the system. This can greatly enhance product choice and flexibility of system configuration for broadcasters.

Two methods exist for adding third-party devices to SIP PBXs. One is to emulate a telephone set, but this can be a complex process. Such phone-like devices will probably need to be *registered* on the PBX, to

alert the central switch via a SIP message that the device is there. Implementation of this also varies in different products. On the other hand, SIP Trunking is generally simple, straightforward and preferred, and is supported by almost all IP PBXs.

GATEWAYS TO/FROM VOIP

Gateways provide the bridge between the SIP/IP network and the telco network. Traditionally, this implied that PSTN, T1 or ISDN was on the telco side. But modern gateways can be used to link to telco-hosted SIP services. In this case, the gateway becomes an enterprise’s IP firewall and router.

For this reason, many gateway devices now have both IP and circuit-switched connections on the telco side. The IP option can be used for both voice and data.

Gateways provide all or a subset of the following:

- Signaling translation from SIP to the telco’s format
- Physical translation between IP and circuit-switched telco networks (e.g., RJ-45 Ethernet to multiple RJ-11s for PSTN connections)
- DTMF¹⁵ tone generation and detection on both the IP and telco sides
- Caller ID detection on the telco side
- Call-progress tone generation and detection (i.e., busy signal, dial tone, etc.)
- Line echo cancellation (digital hybrid)
- Audio transcoding between codecs
- IP router functions
- PBX-like services (not really a gateway function, but often included)

The process of ordering or configuring a gateway will vary depending on the type of interface(s) being used to connect to the telco network. A description of those interfaces follows.

FXS/FXO

These are designations for the two ends of a standard analog PSTN line: *Foreign Exchange Station* (FXS) and *Foreign Exchange Office* (FXO).

An FXS interface emulates a circuit supplied by a telco central office (CO). An FXS supplies talk battery and detects an off-hook condition. It generates 100VAC for a ringing indication. It provides dial tone

¹⁵ *Dual-tone multi-frequency* signaling, used for in-band call routing data on PSTN (also called tone dialing or “TouchTone”).

and other call-progress signals such as ringback and busy. It responds to DTMF tones for dialing and may send caller ID information in modem-encoded audio.

A telephone, and anything that looks like a telephone, is an FXO device. FXO devices signal an off-hook condition by drawing loop current, respond to ringing voltage, provide dialing (either by old-fashioned pulsed loop-interruption or by DTMF) and may detect Caller ID.

T1/E1

These are basic digital interfaces to the switched voice network, and they are in wide use today. This is especially so in the U.S., where T1 is nearly standard for large PBXs. (In Europe, ISDN-PRI is more common for this purpose.) T1 transports up to 24 voice channels, while E1 supports as many as 32. T1s are used in the U.S. and Japan, while E1s are provided by telcos in most of the rest of the world. In addition to the audio, these digital circuits also carry basic signaling in *Channel Associated Signaling* (CAS) bits. This signaling emulates loop-start, ground-start or E&M, depending upon configuration.

T1s can also be used for IP connections. In this case, usually an entire T1's 1.544kbps capacity is used as a transparent pipe from the local IP router to the ISPs equipment. As a result, the term *channelized T1* is now coming into use to distinguish a T1 that is intended for the traditional, circuit-switched voice application described above. A *fractional T1* is a service that uses a portion of the line's full capacity. It is sometimes possible to order a T1 that is divided into a channelized portion and a data-transparent part for IP connectivity.

ISDN-PRI

ISDN-PRI (Integrated Services Digital Network Primary Rate Interface) uses the same underlying circuits as T1s and E1s. Over a T1, 23 speech channels are offered, while an E1 provides 30. One or two of the channels are reserved for signaling communications. This out-of-band protocol transmission allows transfer of information such as calling number, codec type, clearing causes, and such. (Strangely, however, T1 sends Caller ID data via modem-encoding it into the speech channel.) The speech paths are called *B* (*bearer*) channels, while the signaling is carried in *D* (*data*) channels. Almost all large VoIP gateways and PBXs support ISDN-PRI lines. The signaling in the U.S. is a slightly different protocol than that used in Europe and other parts of the world. Gateways will need to be set to match the appropriate protocols used by local telco. Normally in the U.S., this is NI-1 (National ISDN-1), while Europe uses the Euro ISDN standard.

ISDN-BRI

ISDN-BRI (Basic Rate Interface) lines offer two B channels, supported by one D channel. (As noted above, B channels carry audio payload, while D channels carry signaling; this is sometimes referred to as a "2B+D" configuration.) These were intended as a residential replacement for PSTN lines or for small businesses. One application envisaged by its inventors was to allow a simultaneous voice call and data connection. With DSL providing much higher data rates, ISDN-BRIs are moving ever closer to obsolescence.

CONNECTING BROADCAST FACILITIES TO A TELCO VIA VOIP

Although still somewhat exotic at present, telco support for SIP trunking is growing. This kind of connection to telco will reduce and perhaps eventually eliminate enterprise use of T1 and PSTN trunking. If Telcos really do shut down the PSTN, this is what we'll be using to connect our PBXs and on-air interfaces to the outside world.

The physical location of the gateway to the PSTN is inconsequential, as long as the IP path between the service location and the gateway has guaranteed QoS with sufficient bandwidth to support the maximum number of active connections expected. (We know of a California station that has successfully used a SIP trunking provider based in New York state.) In the case that the IP link is to be used for both telephony and data, the system must either have plenty of reserve bandwidth or be designed so that VoIP calls have priority over general traffic. In order to ensure this, there must be only one IP vendor between the service location and the PSTN, and this vendor should guarantee QoS in a Service Level Agreement. When multiple IP service vendors are involved, probability of achieving consistently high quality service and rapid resolution of problems is greatly reduced.

Codec choice is also an issue here. For calls that ultimately are carried by the PSTN, only the native G.711 codec is acceptable for broadcast applications. Use of any other codec would involve a transcoding step, bringing unacceptable reduction in fidelity. This will be especially audible when (traditional) mobile phone calls are involved given their already reduced quality due to low-rate 14.4kbps codecs. Passing those calls through G.711 within the PSTN and then through yet another codec on the way to the broadcast facility over an IP link only makes matters worse.

Full interoperability between the station's VoIP equipment and the carrier's IP service must be verified, as well. Although SIP is a standard, many

vendors enhance it in their implementations with extensions that are not supported by all other vendors.

Something that may be helpful in this area is the *SIPconnect* project from the SIP Forum, a consortium of SIP vendors. The *SIPconnect Interface Specification*¹⁶ was launched by Cbeyond Communications in 2004, with support from Avaya, BroadSoft, Centrepoint Technologies, Cisco, and Mitel. This document details the interconnections between IP-PBXs and VoIP service provider networks. It presents a reference architecture, lists required protocols and features, and suggests implementation rules. It further calls for the G.711 codec to be provided on all equipment and services.

At this writing it is not fully clear whether broadcast facilities should convert their telephone service to SIP-based IP. There is no inherent reason that properly engineered IP trunks would provide anything other than a reliable, high-quality service. Nevertheless, all due diligence is still required by the customer.

MPLS

Multi Protocol Label Switching (MPLS)¹⁷ is an emerging IP service that allows telcos to provide guaranteed QoS when needed by customers, such as for VoIP. MPLS works by adding an MPLS header prefix to IP packets. The prefix contains one or more “labels,” called a *label stack*. These MPLS-labeled packets are able to be switched more efficiently by a Label Lookup/Switch instead of by a lookup on the IP routing table.

MPLS allows class of service (CoS) tagging of packets, and the prioritization of network traffic. Administrators can then specify which applications should move across the network ahead of others. This capability makes an MPLS network useful to enterprises that need to ensure the performance of low-latency applications such as VoIP. Carriers supporting MPLS differ on the number of classes of service they provide and how they price their CoS tiers.

Because it is a standard, however, MPLS may allow QoS to be reliably provided across vendor boundaries, eventually offering QoS to voice applications like the PSTN.

IP Centrex and Hosted PBX Services

Like other IP-based processes, the physical location of a given functionality’s performance is immaterial.

¹⁶ Version 1.0 ratified January 2008; V1.1 is currently in development.

¹⁷ MPLS is defined in IETF’s RFC 3469 (February 2003).

This concept was applied earlier in the discussion of gateways, and it applies to where an enterprise’s IP-PBX is located, as well. This enables *IP Centrex* services or *Hosted PBX* services, in which the hardware is located at the service provider’s site, with no need for such phone system equipment at the enterprise. In a full-fledged installation of this type, a customer’s facility would require only IP phones, which would be plugged into an Ethernet switch, which in turn would connect to the Internet via a router. The primary advantage to such service is that a third party is responsible for installation and maintenance of the back-end equipment. Vendors of these services may also provide a suite of applications that would be difficult to replicate at individual enterprise sites.

SKYPE

Skype is a popular VoIP service provider, and it has been used by some broadcasters for remote origination.

Skype’s technology is proprietary, and its complete workings are therefore not fully understood by the industry at large. It is known, however, that Skype is certainly not SIP-based, so it will not interoperate with other VoIP applications (although it probably uses SIP internally for its SkypeOut and SkypeIn interfaces to PSTN gateways).

For a time Skype used the iSAC codec from the company Global IP Sound (now Global IP Solutions), then used an in-house developed codec with the name SVOPC (Sinusoidal Voice Over Packet Codec). It is a wideband codec with a 16kHz sampling rate, and thus around 7kHz audio bandwidth. A new codec called SILK was introduced in early 2009 in the Skype 4.0 release. It has two modes: 16kHz sample rate with 8kHz audio bandwidth; and 24kHz sample rate with 12 kHz audio bandwidth. It is apparently able to shift between the two modes depending upon network conditions.

Audio streams are encrypted and do not use RTP. Indeed, it seems Skype attempts to obfuscate its streams, perhaps in order to keep firewalls from discovering their presence.

Skype was developed by a group of engineers in Estonia who had developed the KaZaA peer-to-peer file-sharing system. Presumably, Skype uses some technology that was invented during that time. For example, it is generally believed that the user database is stored in a distributed fashion within users’ computers, rather than in a central database.

One of Skype's interesting features is its ability to circumvent firewalls and NATs. Apparently, Skype does this in a particularly stealthy way, which is effective across a wide variety of conditions, but which gives pause to corporate IP managers concerned about security.

Should Skype's popularity continue, broadcast studio systems will have to find a way to elegantly interface to it, perhaps using some kind of server acting as a gateway.

ON-AIR STUDIO PHONE SYSTEMS

While any facility can profit from an IP-based on-air phone system, those using AoIP architectures will achieve substantial additional benefits:

- A single RJ-45 connection from the system Ethernet switch to the on-air phone system interfaces a large number of telco lines and studio audio channels, as well as audio and control signaling to the various user interfaces: telephone-like directors, PCs, and mixing consoles.
- A single on-air phone system server can supply all the studios in the facility with rich telephone capability.
- Each call can have its own hybrid and audio processing. Any number of outputs to console faders can be provided at low cost since there is no need for converters, connectors, and cables for each.
- AoIP is inherently bi-directional, so mix-minus is supported without complication and at no incremental cost.
- A common wiring and Ethernet switch infrastructure serves both studio audio and telecom needs.
- On-air call director controllers can be sophisticated devices owing to their connection over IP.
- Call-screening software running on PCs connects over the same network, and can include integrated softphones for the screeners, streamlining operations and reducing costs.
- Mixing-console control surfaces can incorporate phone system controllers that need no additional connection; their signaling simply uses the network connection already there. Rich status information can be displayed either on the phone control module or the console's main screen.
- Recording and playback of DJ + telephone conversations are simplified. PC-based editors

send and receive audio directly over the network using their native Ethernet connections.

User Interfaces

With a 100BaseT Ethernet connection, there is plenty of bandwidth for sophisticated user interfaces. These may be phone-set-like devices, mixing consoles, or PCs.

There can be rich interaction among the devices. For example, descriptive text regarding a caller can be entered in a PC producer application and appear on the phone's LCD display. A mixing console phone module can select lines and assign them to faders. Once assigned, icons near the fader can show line status.



Figure 5: A desktop controller for an IP-based studio system connects via Ethernet. The large LCDs convey caller information entered into a PC-hosted producer application.



Figure 6: Mixing consoles integrate smoothly into an IP-based on-air telephone system. A single RJ-45 connects the control surface to both the mixing engine and telephone system. A telephone controller module allows convenient line selection and fader assignment at the talent operating position.



Figure 7: Producers use a PC application that includes a softphone, a recorder/editor, and traditional call screening functions. Talent can use the same application. Again, a single RJ-45 connects both control and audio signals.

Telephone Audio Processing

IP-based on-air telephone systems benefit from many of the same audio processing functions that traditional broadcast hybrid interfaces do. These include:

- AGC,¹⁸ on both the input (studio audio send) and output (telephone receive audio) paths.
- Audio response shaping on the send audio to improve intelligibility. Without such filtering, studio microphones may put too much low-frequency energy into the telephone line.
- Automatic multi-band EQ on the telephone receive audio to compensate for the wide variety of telephone sets in the field, as well as effects from different phone lines, codecs and other impairments in speech paths.
- A filter to remove hum and noise on receive audio.
- A “ducker” (or “gate”) to dynamically lower the volume of the telephone audio when the host speaks. This serves both an aesthetic and a technical purpose. As to the first, many talk hosts prefer to have control over the conversation and the ducking helps them to achieve that. As to the technical benefit, a ducker improves the effective, or apparent, send-receive isolation, compensating for deficiencies in the core hybrid’s performance .

¹⁸ *Automatic Gain Control*, a volume-leveling function that reduces long-term dynamic range fluctuations.

TIPS ON IMPLEMENTATION

Besides the installation of VoIP components (such as the IP-PBX) or connections (such as SIP Trunking), perhaps the greatest advantages to the broadcast facility are provided by the integration capabilities of VoIP systems.

A practical implementation of this is found in the integration of VoIP into audio mixing consoles, such as those offered in the *Livewire* AoIP format.¹⁹ Such integration allows rich and direct interaction between the console control surface and the telephone system, providing the ability for enhanced interaction with callers and expanded program-production capabilities.

Cost savings are also significant, given that a single VoIP console module can replace several digital phone hybrids, and wiring paths are simplified. Call screener and/or producer stations are also simplified, in that the same PC running call-screening software can act as a softphone and call director for the VoIP phone system.

Consolidated multi-station or other multi-room systems are also streamlined by VoIP implementation. Instead of the traditional need for running multiple trunk lines from a central switching room to each studio, and managing individual paths for different lines and line types (e.g., PSTN vs. ISDN), VoIP systems can operate over the production IP network. Individual lines need not be dedicated to specific facilities, reducing expense for unutilized resources.

Another useful feature is the remote control capability of VoIP systems. The ability for a single engineer or operator to administer an entire multi-studio telephone facility from any place with an Internet connection (even a wireless handheld device) provides substantial flexibility and agility to adapt to fast-changing requirements.

ONGOING DEVELOPMENT

When the connection from a broadcast station’s listeners to its studios eventually evolves to become IP-based from end-to-end, as it is nearly certain to do, the environment will change further. The possibility for higher fidelity caller audio has been mentioned above, but other, more game-changing effects could also occur.

Consider that an unconstrained pathway for data along with voice might allow a talk show’s producer or host

¹⁹ *Livewire* is a standards-based AoIP system developed by the Cleveland, Ohio-based manufacturer Axia Audio, and supported by a growing number of audio and broadcast equipment manufacturers.

to text-chat with a caller prior to going on air, for example via their PC or smartphone. Callers (or guests) could see a countdown timer to when they will be put on air, and/or when their segment-time expires. Broadcast listeners could participate in instant voting via PCs or smartphones, or view ancillary program text/graphics data. With the addition of IP video streaming, callers (or certainly remote guests) could be seen as well as heard via an Internet link.

Such enhancements could strengthen the relationship between listeners and broadcast programs, bringing broadcast content closer to the style of social media and other participatory “Web 2.0” applications that supply users with the appeal of rich interaction.

The first stage of the application of any new technology is to replicate the function of what came before, but once the new platform is in place, creative people invent new and unexpected ways to use it. IP is a powerful and amazing enabler that has already engendered many surprising things. It is inevitable that more are on the way.

CHALLENGES & CONCERNS

Network Quality of Service

AT&T’s petition prompts a note of concern. As long as voice service is treated by the FCC as just another Internet application, and as long as the FCC holds to its view that providers of Internet service should not be regulated according to application or be allowed themselves to treat various applications differently (the definition of “net neutrality” as defined by FCC Chairman Julius Genachowski), then none of the current regulations regarding voice reliability would seem to apply. Will AT&T and other telcos -- er, that is, ISPs -- be required to provide sufficient end-to-end bandwidth that voice will not suffer from drop-outs and poor fidelity? Indeed, under net neutrality, will providing guaranteed quality-of-service for voice be illegal? Will something like 911’s address reporting be supported somehow?

Delay

An advantage of the PSTN is that it has very little delay. VoIP, on the other hand, always has 10s to hundreds of milliseconds of delay. We’ve gotten used to this in the context of mobile phones, but landlines have usually had no noticeable delay. Audio delay in IP networks is primarily a function of packet size and jitter. The longer the packet, the more time it takes to gather up the audio samples, and the greater the delay.

Jitter determines how many packets have to be buffered in the receiver. The buffer must be large enough to include the latest-arriving packets. In VoIP

systems, the buffer is often a user-configuration item that is set by experience. A value is chosen that results in few packets falling over the buffer time limit.

LANs typically produce no significant jitter, so buffers can be as small as two packets. The public Internet is the most challenging due to the potential for lengthy delays, and moreover because these delays are so variable. Therefore, adaptive buffers combined with effective concealment in the codec provide the best strategy to ensure uninterrupted audio in VoIP.

Delay in VoIP networks produces echo – a talker’s voice being returned to the talker due to some source of leakage along the transmission/receive paths. The usual cause is a poor hybrid at the interface of the digital and analog circuits at the far end of a path that includes a PSTN line. Another source of leakage is mechanical coupling between the earpiece and microphone in the far-end telephone handset, or acoustic coupling when a speaker phone is used at the far end. Such a phone needs to have either a ducker or an acoustic echo canceller that can be relied upon to maintain many tens of decibels of send-to-receive isolation. Because VoIP has more delay than analog or circuit-switched digital speech paths, the demand put on the system for low leakage is higher.



Figure 8: Listeners’ annoyance from echo delay on phone calls is a function of both amplitude and time. In the graph above, TELR is the Talker Echo Loudness Rating, the difference between the perceived volume of the caller’s own voice in real time and that of its echo, measured in dB. Like a S/N ratio, a higher TELR indicates a quieter echo. “Annoyance Contours” shown indicate that results falling into the area below the 10% curve (solid line) are unacceptable, and area above the 1% curve (dotted line) are ideal.²⁰

Listener tests have shown that both the volume and the time-delay of an echo interact to produce a certain level of annoyance, as illustrated in Figure 8. Both the longer and the louder an echo, the more annoyance it produces for the talker. Thus, reducing either the length or the volume of the echo (or both) can help. Echo length is a largely a factor of the variable

²⁰ After ITU G.168, *Digital Network Echo Cancellers*.

Internet delays mentioned previously, but echo volume is something that VoIP equipment designers can fairly consistently control. An echo heard at lower volume actually allows somewhat longer delays to be tolerated.

Generally, VoIP system designers expect to achieve at least 35 to 45dB Echo Return Loss (ERL) and thus they target 150ms as the maximum permissible round-trip delay. IP PBX systems designed for operation on LANs would have much lower delay, perhaps in the 50ms range, so such ERL performance will provide excellent echo tolerance.

Note that these concerns are relevant only to VoIP running on routed wide-area networks. The in-house portion of an IP-based studio telephone system would run on a controlled and switched LAN, so there is no concern with QoS. You can be confident that all packets will arrive, and that they will do so quickly and with very little jitter. Ethernet switches offer plentiful bandwidth at low cost.

Echo is not the only reason to keep delay as low as possible; the natural flow of conversation depends upon delay not being too high, as well. The 150ms VoIP target has also been found to be adequate for this aspect.

Dealing with Echo

Echo cancellation is one of the classical functions performed by broadcast digital phone hybrids in their 2-wire to 4-wire audio conversion. When PSTN lines are used in studios, the send and receive audio signals need to be isolated as much as possible. In studio applications, a hybrid interface needs particularly good send-to-receive isolation. When too much of the send audio leaks through the hybrid and appears in the receive-audio signal fed to the phone input on the studio mixing console, a number of unwanted effects can occur, as follows:

- Distortion of the host's voice. The telephone line will change the phase of the send audio before it returns, with varying shifts at different frequencies. The host audio will be subject to tonal coloration as the original and leakage audio are mixed at the console and combine in- and out-of-phase at various frequencies. As a result, the announcer sounds either hollow or tinny, and this effect can vary from call to call (due to differing phase effects from the variable impedance of each phone call's line characteristics).
- Audio feedback can result from the acoustic coupling created when callers must be heard in the studio on an open loudspeaker.

- When lines are conferenced and the gain around the loop of the multiple hybrids is greater than unity, feedback singing will be audible.
- If the leakage is very high, studio operators will not be able to control the relative levels of the local host audio and the caller because the console telephone fader will affect both signals.

These impairments can occur even when a digital telco line is being used, owing to coupling at the far end (where it is likely that 2-wire conditions remain in effect for the last mile). An IP-to-PSTN gateway or the equivalent function within an IP PBX should always have a *line echo canceller* (LEC) as part of its suite of adaptation functions. But it is not always that case that the LEC rises to "broadcast quality." For that reason, an on-air system attached to a VoIP system may need to have an additional "helper" LEC.

Applications Involving Loudspeakers

A common annoyance in broadcast studio operations is the feedback that results from using a loudspeaker in the studio to listen to telephone calls (typically done to avoid talk show guests in the studio from having to wear headphones). This arises from the acoustic coupling of the sound emanating from the loudspeaker into the studio microphone. Ducking helps by reducing the gain "around the loop," but it compromises full-duplex operation and can cause problems for the caller and host hearing each other.

The earlier reference to a ducker only considered its insertion in the telephone-audio *receive* path. In order for a ducker to help with feedback, it also would need to act in complementary fashion on the studio-audio send path. There is also the fundamental problem that any acoustical reverberation from the studio would be heard by the caller. Thus when callers talk, their voices bounce around the studio and are sent back through the studio microphones with additional reverb "tails" from room reflections. Adding this time-dispersion to the round-trip transmission delay can be very distracting to the caller.

In such situations, *Acoustic Echo Canceling* (AEC) provides a solution. AEC removes the caller's loudspeaker audio from the studio microphone signal, leaving only the host voice in the send audio returned to the caller (see Figure 8). AEC has been used in high-end audio and video conferencing systems for many years. Higher quality broadcast hybrids and on-air telephone systems also have included a limited form of AEC for some time. But only recently has AEC technology advanced to the stage where it is both truly effective and affordable, thanks both to breakthroughs in the design of adaptive AEC algorithms and the ever-increasing power and lower-

cost of processor chips. This has serendipitously occurred at a time when the additional delay of mobile and VoIP connections make AEC nearly essential for broadcast-studio telephone audio.

These latest-generation AECs are quite effective, allowing substantial attenuation of even very high-volume caller audio from studio loudspeakers in the resulting send audio signal. Unlike previous designs, these AECs operate at up to 20kHz audio bandwidth, so they are ready for emerging wideband VoIP codecs. Another improvement over earlier systems is their ability to dynamically adapt to changes in acoustical conditions during a call (such as studio microphones being opened, closed or moved). Earlier “time-domain” AECs depended upon the acoustic path remaining fixed, and could quickly degenerate into feedback when conditions changed during a call. The frequency-domain technology used by newer AEC equipment can dynamically adapt to changes in the reverberant field as picked up by studio microphone(s)

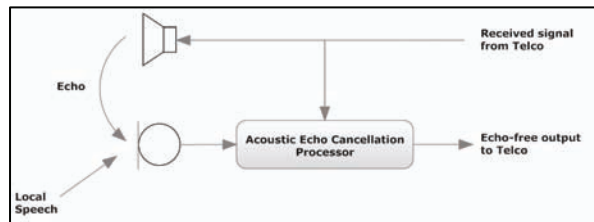


Figure 9: An Acoustic Echo Canceller is required for smooth full-duplex conversation when a loudspeaker is used for hearing telephone audio in the studio.

This new AEC technology is particularly useful for TV studio applications where it can be impractical to have talk show guests using headphones, or even earbuds. These programs also like to use roving hosts with handheld microphones. Today’s high-performance AEC technology allows talent to move around while the guests and audience listen to phone calls on studio loudspeakers.

REFERENCES

Comments – NBP Public Notice #25: Comments Of AT&T Inc. On The Transition From The Legacy Circuit-Switched Network To Broadband, 12/21/2009.

Church, Steve and Skip Pizzi. Audio Over IP: Building Pro AoIP Systems With Livewire. Burlington, MA: Focal Press, 2010.

Alexander, J., et al. Cisco Call Manager Fundamentals. 2nd Edition. Indianapolis: Cisco Press, 2006.

Bormann, C. (Editor), et al. “RObust Header Compression (ROHC): Framework and four profiles:

RTP, UDP, ESP, and uncompressed.” RFC 3095. IETF, July 2001.

Camarillo, Gonzalo. SIP Demystified. New York: McGraw-Hill, 2002.

Church, Steve and Rolf Taylor. “Telephone Network Interfacing.” NAB Engineering Handbook, 10th Edition (2007): 609-644.

Davidson, Jonathan and James Peters. Voice over IP Fundamentals. Indianapolis: Cisco Press, 2000.

Dierks, T. and E. Rescorla. “The Transport Layer Security (TLS) Protocol, V1.2.” RFC 5246. IETF, August 2008.

H. Schulzrinne, et al. “RTP: A Transport Protocol for Real-Time Applications.” RFC 3550. IETF, July 2003.

Handley, M. and V. Jacobson. “SDP: Session Description Protocol.” RFC 2327. IETF, April 1998.

Handley, M., et al. “SIP: Session Initiation Protocol.” RFC 2543. IETF, March 1990.

Handley, M., V. Jacobson and C. Perkins. “SDP: Session Description Protocol.” RFC 4566. IETF, July 2006.

Hersent, Oliver, David Gurle and Jean-Pierre Petit. IP Telephony: Packet-Based Multimedia Communications Systems. Reading, MA: Addison-Wesley, 1999.

International Organization for Standardization (ISO/IEC). “Low Delay AAC Profile.” ISO/IEC 14496-3:2005/Amd 1:2007. March 2005 (Amended January 2007).

International Telecommunications Union, Telecommunications Sector. “40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM).” ITU-T Recommendation G.726. December 1990.

—. “7 kHz Audio-coding Within 64 kbit/s.” ITU-T Recommendation G.722. November 1988.

—. “Coding of Speech at 8 kbit/s Using Conjugate-Structure Algebraic-Code-Excited Linear Prediction (CS-ACELP).” ITU-T Recommendation G.729. January 2007.

—. “Digital Network Echo Cancellers.” ITU-T Recommendation G.168. March 2009.

—. “Dual Rate Speech Coder for Multimedia Communications Transmitting at 5.3 and 6.3 kbit/s.” ITU-T Recommendation G.723.1. May 2006.

—. “G.729 Based Embedded Variable Bit-rate Coder: An 8-32 kbit/s Scalable Wideband Coder Bitstream Interoperable with G.729.” ITU-T Recommendation G.729.1. May 2006.

—. “Low-Complexity Coding At 24 And 32 kbit/s for Hands-Free Operation in Systems with Low Frame Loss.” ITU-T Recommendation G.722.1. May 2005.

—. “Pulse Code Modulation (PCM) of Voice Frequencies.” ITU-T Recommendation G.711. November 1988.

—. “The International Public Telecommunication Numbering Plan.” ITU-T Recommendation E.164. February 2005.

—. “Wideband Coding of Speech at Around 16 kbit/s Using Adaptive Multi-Rate Wideband (AMR-WB).” ITU-T Recommendation G.722.2. July 2003.

—. “Wideband embedded extension for G.711 pulse code modulation.” ITU Recommendation G.711.1. March 2008.

Pelletier, G. and K. Sandlund. “RObust Header Compression Version 2 (ROHCv2): Profiles for RTP, UDP, IP, ESP and UDP-Lite.” RFC 5225. IETF, April 2008.

Rosenberg, J., et al. “SIP: Session Initiation Protocol (SIP v2).” RFC 3261. IETF, June 2002.

Schulzrinne, H., et al. “RTP: A Transport Protocol for Real-Time Applications .” RFC 1889. IETF, January 1996.

Sharma, V. and F. Hellstrand. “Framework for Multi-Protocol Label Switching (MPLS)-based Recovery .” RFC 3469. IETF, February 2003.

Sibley, C. and C. Gatch. “IP PBX/Service Provider Interoperability.” SIPconnect 1.0 Technical Recommendation. 2008.

Sinnreich, Henry and Alan B. Johnston. Internet Communications Using SIP. New York: John Wiley & Sons, 2001.