TEXAS INSTRUMENTS

Building Residential VoIP Gateways: A Tutorial

by T Y Chan and Debbie Greenstreet,
Glenn Yancey and Kim Devlin-Allen,
David Jarrett and Keith Buchanan,
Sophia Scoggins PhD,
Matt Harvill and Demetri Jobson
VoIP Business Unit

as published in...

analog ZONE

www.analogzone.com

Building Residential VoIP Gateways: A Tutorial

Part One: A Systems-Level Overview

by T Y Chan and Debbie Greenstreet,
VoIP Group,
Texas Instruments Incorporated

While voice-over-IP (VoIP) products have been deployed in the market for over seven years, recent announcements by service providers such as Vonage, AT&T, Sprint and others have created a flurry of activity by consumer equipment manufacturers racing to roll out residential VoIP gateway products. These low-cost devices are usually standalone boxes that provide VoIP functionality for POTS (plain old telephone system) via a broadband modem (usually cable or DSL). They serve as a bridge between the TMD/analog POTS world, and the IP-centric, packet-based world of the Internet.

As with most consumer products, their designers are usually faced with meeting aggressive product cost targets along with tight development schedules. The product feature shopping list often includes features not only specific to the basic VoIP gateway functionality but to other ancillary functions as well. These include data bridging and routings, such as found in common residential router products, emerging voice and signaling security features such as voice encryption and IPSec, and quality of service (QoS) features necessary to troubleshoot and maintain residential VoIP services.

This article is the first in a series intended to assist engineers by providing detailed design considerations for all major portions of VoIP residential gateway products. This part serves as a functional overview of the major components often required in today's VoIP residential gateways. It goes into detail about the key elements necessary for a quality voice over IP call, and highlights some of the design considerations of the telephony circuitry (which will be explored in additional detail in a subsequent article).

Overviews of the security, data routing and QoS monitoring elements are provided as well, to serve as an introduction to more detailed design analyses that will follow in subsequent articles.

Mimicking POTS

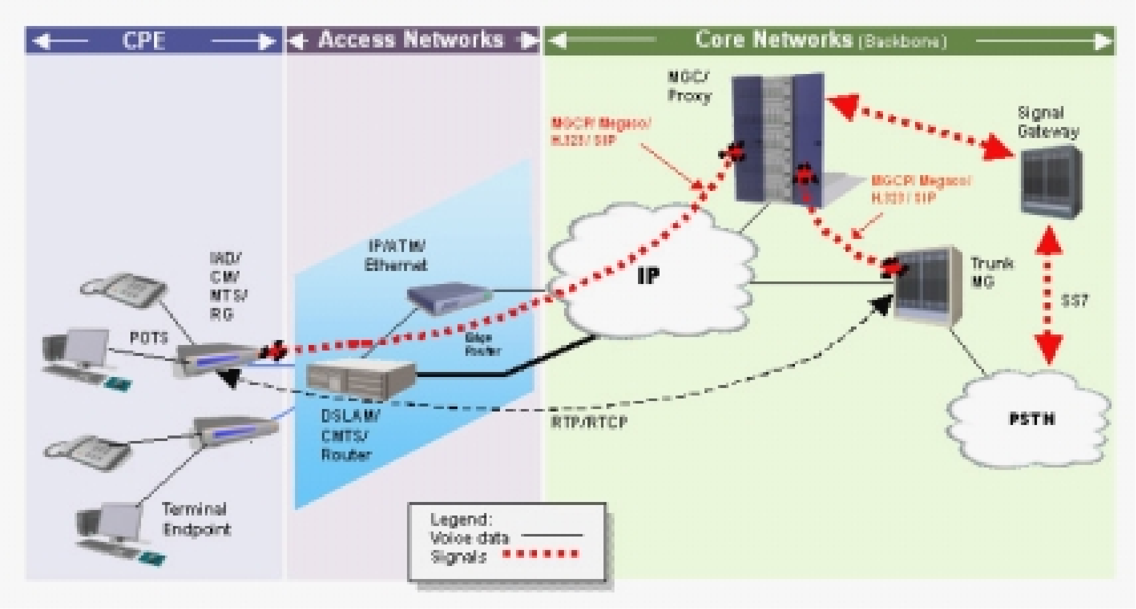


Fig. 1: Network's-Eye View Of A VoIP Call

Since the majority of today's installed base of handsets are still POTS-type units, most Voice over IP calls originating from customers premise equipment (CPE), come from a POTS phone connected to a voice gateway. The unit at the remote location could be another VoIP CPE device or simply a POTS phone connected to the PSTN. The flow of a sample VoIP call through the network is depicted in Fig. 1.

Of course, for any VoIP system to be successful, it must first provide the equivalent experience that end users have grown accustomed to with current POTS systems. To replicate the user's traditional interface with the PSTN, tone generation and detection functions are necessary. Dialed digits must be accurately collected and replayed at the receiving end to successfully execute a call.

Tone generation/detection is not only necessary at the beginning of the call. Tone-driven features like voicemail and calling card functions, tone generation/detection must be handled mid-call as well. Therefore the VoIP processing must also support the ability to successfully transmit dual-tone multiple-frequency (DTMF) tones in-band; however, vocoders with compression may distort these tones. To avoid these potential distortions, designers need to turn to advanced techniques such as IETF request for comment RFC2833 when passing tones in conjunction with the use of low bit rate vocoders. A tone generation function is necessary to provide depressed tone playback and call progress tones. The ability to detect tones and properly switch processing for fax and data modem signals is also a requirement, as all types of telephony currently available on the PSTN must be supported.

Dealing with Echo

Combating problems with echo are an essential element to VoIP adoption in the traditional telco world. As VoIP moves to replace the PSTN system, it must adopt robust echo-cancellation techniques to meet the demands of packet networks.

Echo is present in conventional POTS networks, and the PSTN employs echo cancellers throughout the system. Line echo is caused when a connection involves conversion between a four-line to a two-line telephony hybrid. Echo is generated toward the packet network from the telephone network.

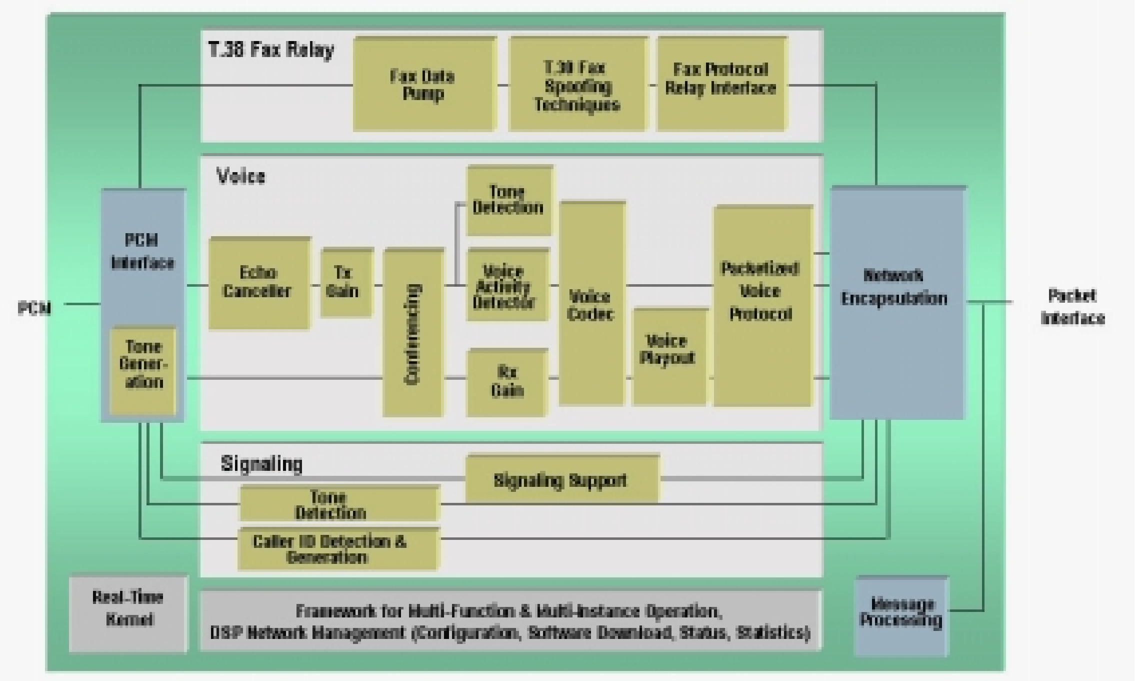


Fig. 2: Block Diagram Of A TDM-IP Gateway

PSTN specifications dictate that echo cancellation functionality is necessary when the delay exceeds 50 ms. However, because the IP network portion of the VoIP solution almost always adds more than 50 ms of round trip delay, line echo cancellation is essential when VoIP solutions interface to the PSTN. To do this, the echo canceller, as shown in Fig. 2, compares the voice data received from the packet network with voice data being transmitted to the packet network. The echo from the telephone network is removed by a digital filter on the transmit path into the packet network.

The echo-cancellation tail length, that is the length of the echo required to be cancelled by the processor, varies among different VoIP applications. The tail-length requirement is determined by the distance between the gateway equipment (residential gateway) and the four-to-two line hybrid. Typically this ranges from an 8-ms tail size for residential/SOHO applications to 32 ms, compared to up to 128-ms tail sizes for carrier applications.

Since most phone calls established via residential VoIP gateways will, at some point, terminate to PSTN equipment, line echo cancellation is required. For residential gateways, a typical length of 8- to 16-ms echo-cancellation tail-length capability is usually sufficient. As a minimum quality benchmark, the echo-cancellation functionality should be compliant to ITU G.165 and G.168 standards.

Voice Encoding

Voice encoding is necessary to convert the analog signal to voice packets. This often includes compression to reduce the 64-kbit/s stream produced by G.711-PCM-encoding stream (used by most traditional PSTN trunk lines) to a lower bit rate for more efficient transport across both the network and the subscriber's "last mile" link.

Typical vocoders used in VoIP systems today include G.729ab and G.723.1. The G.729ab vocoder offers data rates as low as 8kbit/s and the G.723.1 at 5.3 and 6.3 kbit/s. The tradeoff between these low bit rate vocoders and G.711 is reduced bandwidth utilization vs. slightly higher voice quality. The G.729a is an optimized implementation of the very common G.729 voice compression algorithm. It is important to note that G.729 is the base algorithm and that G.729 is interoperable with G.729a. G.729ab, the appendix B portion of this algorithm, incorporates the voice activity detection function in the vocoder itself.

Detecting Voice

Voice activity detection (VAD) and related silence suppression, whether incorporated in the codec or as an external software function, should also be supported as a configurable (enable/disable) feature in VoIP designs. The VAD monitors the received signal for voice activity. When no activity is detected for a specific period of time the software prevents unnecessary packetization and subsequent transmission of silence. This function also measures the idle noise characteristics of the telephony interface and noise measurements are subsequently relayed to the receiving gateway. Comfort noise generation (CNG), the playout of low-level background noise to the receiver, is recommended for user confidence in the call connection. If the call appears too quiet, users may anticipate that the call has been disconnected.

Residential gateways must also support the use of fax relay techniques. Fax relay offers bandwidth reduction and a more robust, reliable means of connecting fax over IP calls, and is a very popular feature for SOHO and SMB equipment. Fax relay functionality involves demodulation of the facsimile scan data, encapsulation into IP packets, and subsequent demodulation of the fax IP packets at the receiving gateway. This requires support of the T.30 fax protocol implemented between the fax machine and the VoIP gateway, as well as T.38 fax IP packet encapsulation for IP transmission.

POTS interfaces

Residential VoIP gateways interface to traditional telephony equipment through FXS signaling to a pulse code modulation (PCM) interface. This interface receives PCM samples from an analog codec interface and forwards them to the appropriate functions, such as those described above. Conversely, the interface forwards processed PCM samples received from the DSP to the digital interface. The PCM interface performs continuous re-sampling of output samples to avoid sample slips.

Playout

When voice and fax samples have been processed, they must be packetized. VoIP systems typically employ real-time packets (RTPs). On the receive side a voice playout unit is necessary to buffer received voice packets and forward them to the vocoder for playout to the user. This playout unit also serves as a jitter buffer/manager to queue several packets, avoiding packet under-run or over-run.

Implementing The Features

The features described above are typically implemented in software, usually on a DSP. They can, however, be implemented in a RISC processor. This is advisable only when the additional processing requirements for the typical RISC functions are minimal. For RISC-only VoIP architectures, the available CPU cycles (MIPS) must be carefully managed between voice processing and network/telephony signal processing. The functions described so far represent the telephony signal processing tasks required for supporting VoIP media streams. The VoIP gateway must also support signaling protocols, for both the telephony side and the packet side of the gateway.

Packet And Telephony Network Signaling

Translation of the telephony signals to packets is only a part of the VoIP gateway solution. A gateway must also support telephony control signals, such as on-hook and off-hook functions as well as network control signals or protocols such as Session Initiated Protocol (SIP) -- both of which place very different demands on the host processor and its software. There is a comprehensive set of processing tasks that are typically executed on a RISC processor that translates the telephony signals/protocols to the packet protocols and vice versa.

Analog Phone And PSTN interfaces

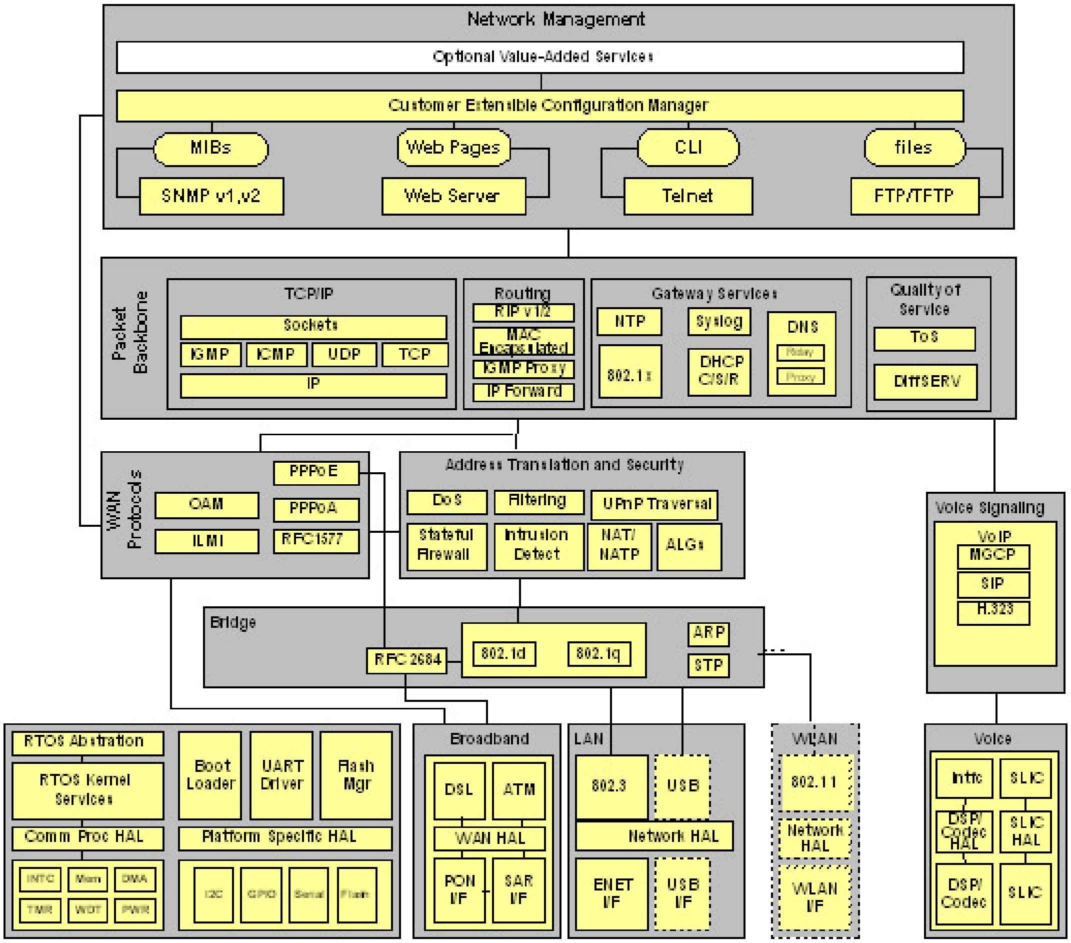


Fig. 3: Software Functions In A Typical VoIP Gateway

In most applications, a software-based state machine-like service that serves as a call controller handles these functions (See Fig. 3). The call control executes the necessary functions through a stage of each call in the gateway. Another critical element of the signaling process is the network protocol stack itself. SIP is a very popular protocol for the residential gateway market; however there are some deployments of Media Gateway Control Protocol (MGCP) and H.323.

Supplementary Services And Device Provisioning

VoIP gateways used in residential applications require the support of functions typically available in phone services today. This includes features such as call waiting, call forwarding, visual message waiting indicator and call transfer. Software is required to interpret these commands from the network and execute the function through the gateway to the telephone.

As a remote device in the service provider network, the residential gateway must be able to be configured either on premises or remotely, but not require separate monitors or other equipment. This configuration requires software in the gateway to accept and process the provisioning and it is desirable that the user interface be simple and easy to use. This provisioning software is not insignificant. There is also a preference in the market that residential gateway devices have the capacity for dual image (program load) storage, such that a program update can be downloaded without deleting the current image. This has impact on the overall software program design, as well as FLASH and SDRAM requirements.

Fig. 3, again, shows a comprehensive view of the functional blocks required in a complete VoIP residential gateway system. In addition, device drivers, Ethernet interfaces, real time operating systems (RTOS), and IP stacks should not be overlooked.

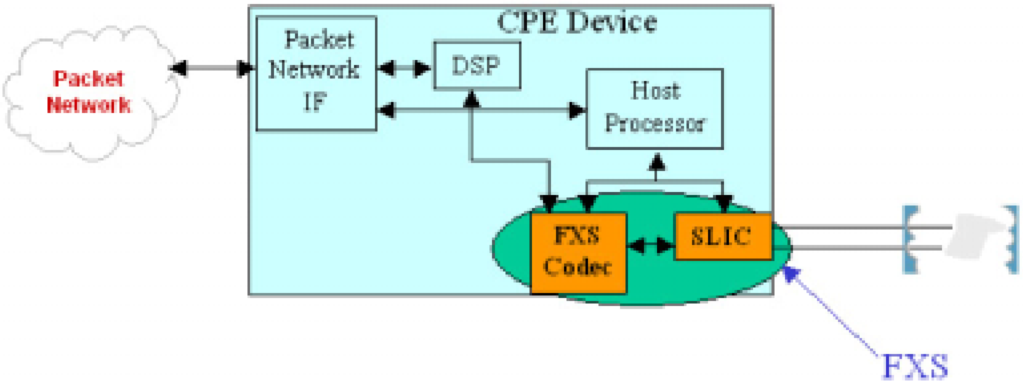


Fig. 4: A Foreign Exchange Station (FXS) Supporting A POTS Phone

To connect the old reliable analog phone to the voice gateway, a Foreign Exchange Station (FXS) is needed (see Fig. 4). In some applications, for outside calling using the PSTN, an FXO connection is needed (see later Fig. 5).

The FXS consist of two parts, a codec and a SLIC. A codec is comprised of an ADC and a DAC. The ADC is used convert the analog signal from the analog phone into digital signal for transmission onto the VoIP network. The DAC is to convert digital signals to analog levels to drive the analog phone. The sampling rate for ADC and DAC is usually in the 8-kHz range in order to achieve an audio bandwidth of 4 kHz. The Subscriber Line IC (SLIC) device emulates PSTN networks' voltage levels. It needs to detect on-hook, off-hook and generate ringing voltages which can range to 120 V. Its main function is to combine the analog signal with the PSTN voltages.

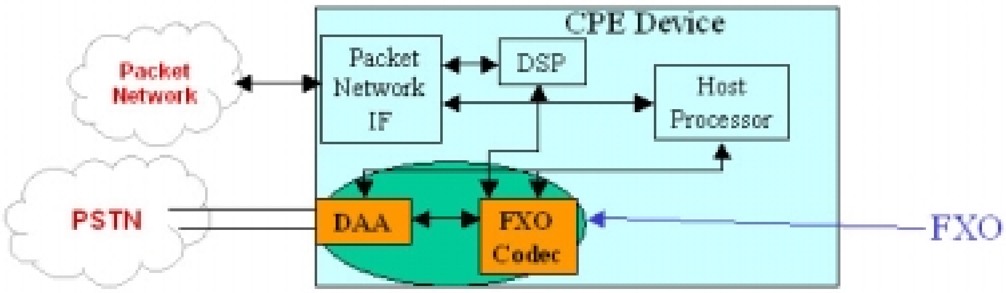


Fig. 5: A Foreign Exchange Office (FXO) In A Residential Gateway

For the case for when a voice gateway CPE device needs to connect to a local phone company, this requires a Foreign Exchange Office (FXO) interface (fig. 5). The FXO consists of a codec and the data access arrangement (DAA). The codec has the same functionality as in the FXS, while the DAA emulates a (POTS) phone. Its main function is to remove the high voltage dc bias, and it only passes the analog ac signal through when coming from the PSTN system, by applying a loop closure towards the PSTN.

Uses for FXO ports are:

- Lifeline for power failure: used when there is no power to the voice gateway, which prevents calls from connecting to the packet network. In this case, the analog phone (through FXS) connects directly to the FXO port through a relay
- Call redirection: used in the case when the subscriber dials a number that is unreachable through the packet network. In order to complete the call, the voice gateway will redirect the call through the FXO port. For a user-friendly gateway the CPE device dials the digit to the FXO port, which prevents customers from having to redial the numbers
- Remote VoIP calling: when a customer is not at home, they can still make a VoIP call by calling their home number through the PSTN network. The voice gateway picks up the call through the FXO port, and forwards it to the VoIP network.

Additional details on FXS and FXO design circuitry will be discussed in upcoming articles.

Voice Gateway Data Functions

In CPE residential applications, the voice gateway is normally connected on the LAN side of a broadband modem. If the household has more than one PC, the voice gateway can be a standalone device terminating IP connections by connecting to a router or hub. If the home has a single PC, then introducing a voice gateway will involve creating a home network and purchasing a router or hub. To ease the adoption of packet voice services, the most appropriate configuration for a voice gateway is to include a data routing function for connecting another PC. This way, the PC connects to the LAN side and the modem connects to the WAN side of the broadband modem. In this type of configuration, the voice gateway should include data routing functionality.

In deciding on the functionality and performance of data functions, it is useful to understand the application and configuration of the broadband connection. With the exception of VDSL, most residential broadband modems have capacities well below 50 Mbit/s. Therefore, in designing a voice gateway, it is necessary to understand the end-user application in order to determine the appropriate price/performance goals. Some of the data functions that should be included are:

- Routing
- NAT, NAPT, dynamic and static
- Firewall
- DHCP client / server
- PPPoE
- TFTP

Including these popular and useful functions makes the transition to VoIP easier for the consumer to connect with their broadband service. Further, incorporating a voice gateway into a router or hub lowers the cost of ownership by not requiring customer to purchase a separate box. When the household purchases another PC, a switch can be purchased to network the PCs and voice gateway together.

VoIP Security Elements

Secure voice communications is receiving a great deal of attention by service providers deploying residential VoIP services. Secure VoIP implementations can leverage many security elements already established for data communications. One of the key functions of the current Internet security infrastructure is monitoring the integrity of the data transmitted. This element covers both the assurance that the message between two entities has not been tampered with, as well as the authentication of the recipient. A similar element is the support for non-repudiation, which is the rejection of a digitally-signed message (by secure keys). The confidentiality level of Internet security ensures that the recipient and the transmitter of the message are the only ones that may view the contents of such a message. The authorization function of the security element suite assures a network user access to a particular network service only upon verifying identity.

Depending upon the level of security concern by end users or service providers, various levels of security features may be required. One common feature is encryption of the voice payload itself. Another level of security might require encryption of the signaling messages that establish the phone call.

Pulling it All Together

While there is pressure to put together a residential gateway solution at the lowest possible cost, it imperative that the components selected achieve optimal quality and performance. The voice processing, network and telephony signaling, POTS interface and Ethernet interface are the minimum functions required to develop these gateways. It is also essential to understand the types of supplementary services and the extent of provisioning functions required in order to ensure that the product is complete. Regional considerations and programmability requirements will dictate the type, and ultimately the cost of the POTS interface. In addition, for residential gateways requiring advanced features such as data-routing functionality or voice encryption or authentication, require additional processing power. If including these features, a designer must take care to ensure that the proper amount of processing power and a sufficiently-flexible architecture is available to support such requirements. Subsequent articles in this series will go into further detail on the issues and tradeoffs of these features.

About The Authors

Debbie Greenstreet dgreenstreet@ti.com is the product management director for TI's voice-over-packet group. She is responsible for product definition and direction of voice-over-cable and SME voice gateway products. Debbie has more than 18 years of experience in the networking and telecommunications field, in hardware and software design, as well as program and product management at companies such as Hyundai Network Systems and Raytheon. She earned a BEE at the University of Virginia.

T Y Chan tychan@ti.com is a senior technical staff member with TI's voice-over-packet group, serving as the system engineering manager for VoIP customer premise equipment products. Since joining TI in 1989, T Y has held various positions in the company, including engineering manager for DVD solutions and system manager for PC processors. He earned both a BSEE and MSEE at Bradley University.

Part Two: VoIP Telephony Interfaces

By Glenn Yancey and Kim Devlin-Allen,
VoIP Group,
Texas Instruments Incorporated

When developing a VoIP system, one key area of consideration is the interface to an analog telephone. The designer must understand the telephony requirements that exist in the PSTN, as they must also be supported in VoIP systems. These articles are intended to provide engineers with design considerations for all major portions of their VoIP product. In this portion, we'll focus on the two most common interfaces to a standard POTS phone: Foreign eXchange Subscriber (FXS) and Foreign eXchange Office (FXO). It will describe the functionality provided by FXS and FXO circuits, cover some history of FXS and FXO, discuss industry standards, and highlight some of the challenges designers may face when supporting analog telephony interfaces on their VoIP residential gateway.

FXS and FXO are common terms in the world of analog telephony, but what is the difference between the two and why are they important in VoIP applications? In a traditional telephone connection over the PSTN, the telephone central office switch feeds battery and provides ringing to the phone. The phone itself completes the tip/ring circuit to request service or answer a call from the PSTN. For calls placed over the Internet, the FXS circuit emulates the telephone central office switch. The residential gateway "pretends" to be the switch, providing both battery and ringing to the phone and detecting loop current. The FXO circuit, on the other hand, emulates a phone, providing loop closure and detecting incoming ringing.

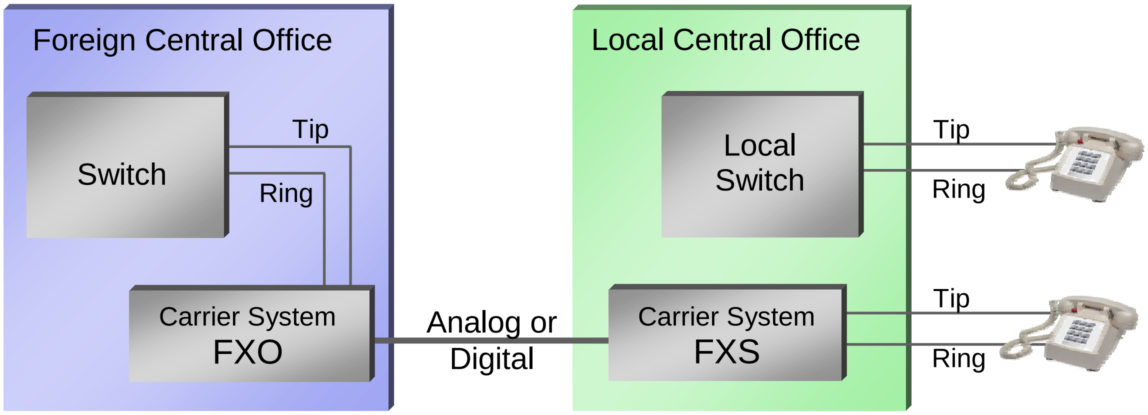


Fig. 1: Dial Tone From A Foreign Central Office

The terminology of FXS and FXO came from the desire to enlarge local calling areas. Before 800 toll-free calling was available, business customers seeking alternatives to expensive long distance charges were offered a foreign dial tone service. Carrier systems, first analog and then digital, were created to support this service, extending dial tone from a foreign central office (**F**oreign e**X**change **O**ffice) to multiple local central office sites

(**F**oreign e**X**change **S**tations). This application was one of the earlier uses for the FXO interface and is responsible for the terminology that still exists today.

Analog Phone And PSTN interfaces

The FXS circuit consists of two parts, a CODEC and a SLIC (Subscriber Line Interface Circuit). A CODEC is comprised of an ADC and a DAC. The ADC converts the analog signal coming from the analog phone into a digital signal for transmission over the VoIP network. The DAC converts digital signals to analog levels to drive the analog phone. In order to achieve an audio bandwidth of 4 kHz, the sampling rate for both the ADC and DAC is usually around 8kHz. The SLIC device emulates PSTN voltage levels. It must detect if the phone is on-hook or off-hook and generate ringing voltages up to 120 V.

The circuitry for the FXO consists of a CODEC and a data access arrangement (DAA). The CODEC has the same functionality as in FXS, converting analog speech to digital signals, and vice versa. The DAA emulates a (POTS) phone. Its main function is to remove the high voltage dc bias, passing only the analog ac signal from the PSTN by applying a loop closure towards the PSTN.

FXO Mirrors FXS

In VoIP gateways the FXS circuit is the primary interface for establishing outgoing calls and receiving incoming calls over the packet network. In a central office application the two-wire SLIC interface on a POTS line card serves as the FXS interface. In CPE applications, the FXS circuitry exists in the gateway, providing dial tone, battery current and ring voltages and detecting loop closure from the phone. Because this switch functionality resides at the CPE level, a direct connection to the PSTN is not necessary. There are cases, however, when a connection to the PSTN is useful using the FXO interface. It presents the same type of interface to the central office as an ordinary POTS telephone, with some improvements. Some important uses of the FXO port include:

- Lifeline for power failure: when there is no power to the voice gateway, the gateway is not able to connect to the packet network to place or receive a call. In this case, a relay can be used to connect the analog phone directly to the PSTN. When this situation occurs, the FXO circuit is intelligent and can detect a call is in progress, preventing that call from being disconnected once power is restored.
- Call re-direction: when the packet network is unavailable due to network congestion, the FXO circuit can remember the number dialed by the subscriber and route the call through the FXO circuit to the PSTN, to complete the call. This process prevents customers from having to redial the phone number when the packet network is down.
- Remote VoIP calling: when a VoIP customer is not at home, they can still make a VoIP call by calling their home number through the PSTN network. The voice gateway receives the call through the FXO port and forwards it to the VoIP network.

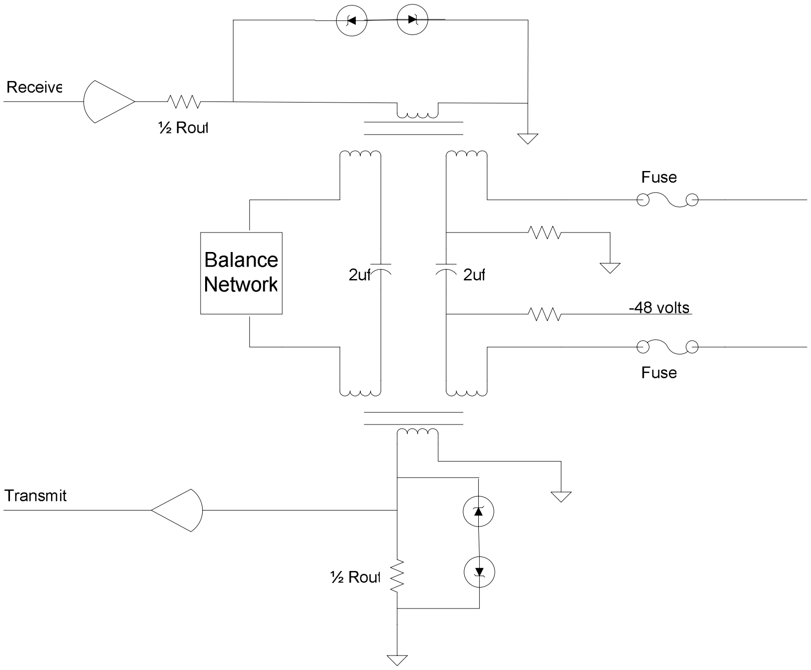


Fig. 2: Basic FXS Interface

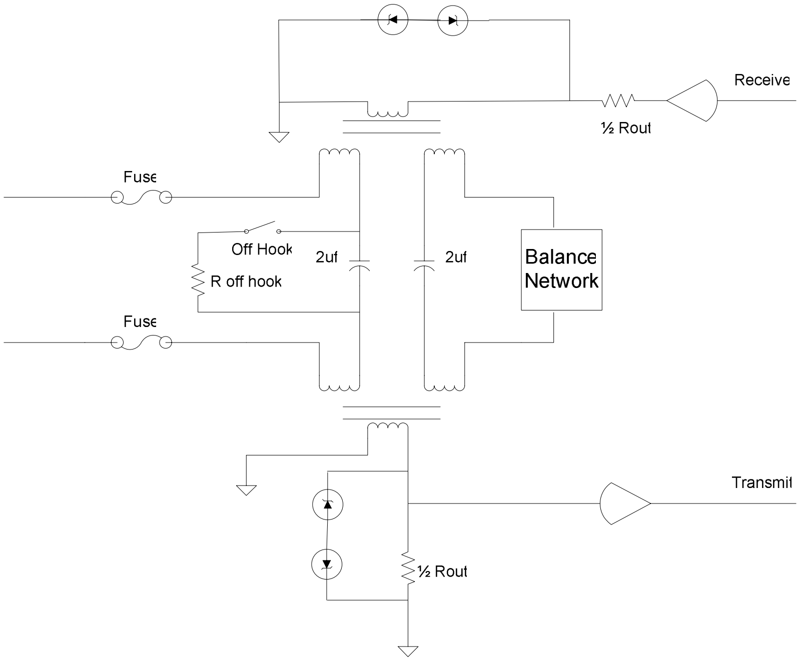


Fig. 3: Basic FXO Interface

Figs. 2 and 3 show how the FXS and FXO interfaces provide some common functions, referred to as "BORSCHT" functions. The BORSCHT terminology is oriented towards FXS functionality, while FXO tends to be a mirror image of some of these functions.

B: Battery feed function found in an FXS linefeed interface. The complementary function in an FXO interface is battery sink. As seen in Fig. 3 a connection is made between the central office tip and ring leads by the FXO off-hook relay, with current limiting being provided by the FXO.
O: Over-voltage protection must be provided by the FXO due to exposure to lightning and power cross conditions. The SLIC's Tip and Ring inputs in the FXS circuit are designed to provide additional over-voltage protection.
R: Ringing is provided by the central office, but the FXO must be able to detect ringing and forward this information. The FXS circuit must provide ringing to the phone. A low-voltage ring signal generated by either the CODEC or SLIC is amplified by the SLIC and placed on the local loop to ring the phone.
S: Signaling refers to the ability of the FXO to receive on/off-hook information and present an off-hook on-command to the central office. It must also detect ringing and other conditions and transmit this information. The FXS must be able to detect on/off-hook states, detect and generate DTMF tones, and generate signals for caller ID.
C: Coding is a function of the CODEC devices that are part of both the FXS and FXO interfaces. It refers to the A-to-D and D-to-A coding of the voice signal.
H: Hybrid functionality is essential for stability and good voice quality and is equally important in both the FXS and FXO interfaces. Echo functions are detailed below.
T: Test is not normally an FXO function as automated maintenance and testing is provided by the central office. However, because the FXS circuit bypasses the PSTN, the required test and diagnostic functionality are included in the CODEC/SLIC.

Echo

The importance of stability and good voice quality are essential whether a call is made over the PSTN or packet network. The potential impact of echo is critical to the functions of both the FXS and FXO interfaces. Note too, that special hybrid functionality is required in both cases within the CPE device to handle the various line impedances in the world. Ordinary POTS telephones have relatively uncontrolled impedances between 200 Ω and 400 Ω. Since the current from office to subscriber is two-wire with no gain added, the impedance variations typically encountered do not affect performance. Stability and line echo issues can arise, however, when a carrier system uses two-to-four wire voice frequency (VF) hybrids on each end, as well as possible gain in the four-wire path.

Line echo results from either the delayed "bleed-though" of the transmitted voice signal into the receive path at the hybrid (2-wire to 4-wire conversion point) or from reflections in the local loop due to impedance mismatches. Line echo is always present in the PSTN and is not necessarily a problem. In fact, some of your telephone's transmit signal is coupled into the receive path in order to generate sidetone. Sidetone lets the speaker hear his or her own voice in the receiver. Without sidetone, the speaker would be unsure if he or she was being heard on the other end, and could make for an awkward conversation.

When uncontrolled, however, excessive line echo can affect a caller's experience in two ways:

- The louder the echo, the more disruptive it will be during a voice call. Many times low levels of echo are present on the line, although they are not detectable by the user.
- The length of delay of the echo also greatly affects voice quality. This delay represents the time that elapses between when the user speaks, and when he or she hears their echo. Round trip echo delays greater than 25 ms will begin to affect voice quality.

The main function of the hybrid circuit that completes the 2-to-4-wire conversion and vice versa is to limit the amount of outgoing transmit signal that "bleeds" into the incoming receive path. As a result of transhybrid imbalance (hybrid component imperfections, impedance mismatches, etc.), some amount of Tx signal always gets into the Rx path (see Fig. 4).

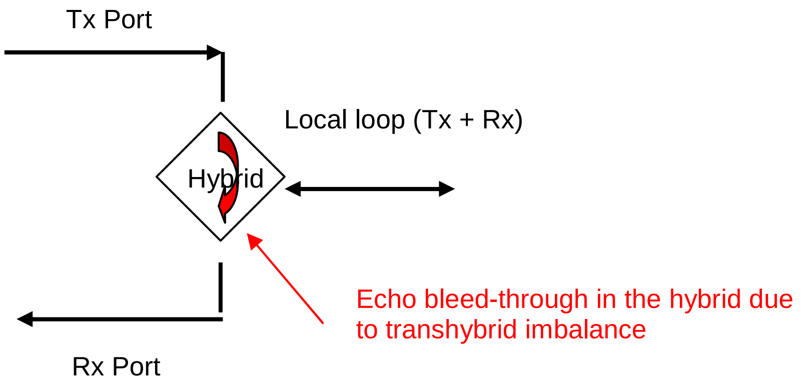

Tx Port
Local loop (Tx + Rx)
Hybrid
Echo bleed-through in the hybrid due to transhybrid imbalance
Rx Port

Fig. 4: Line Echo At 2-To-4-Wire Conversion

In addition to the echo caused by transhybrid imbalance, hybrid termination impedance mismatches can also cause line echo. If a line is not correctly terminated with its characteristic impedance, echoes will be generated. This echo is a result of the incoming signal from the 2-wire local loop hitting the hybrid termination resistance and reflecting back down the line (Fig. 5). Improperly terminated CPE equipment such as phones or modems can also generate echoes in the local loop.

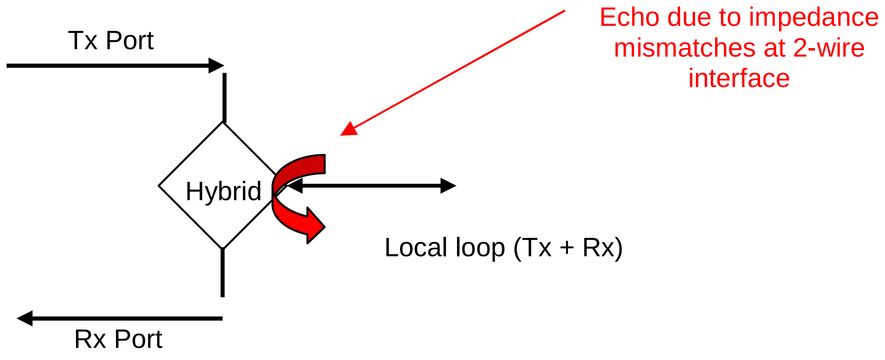
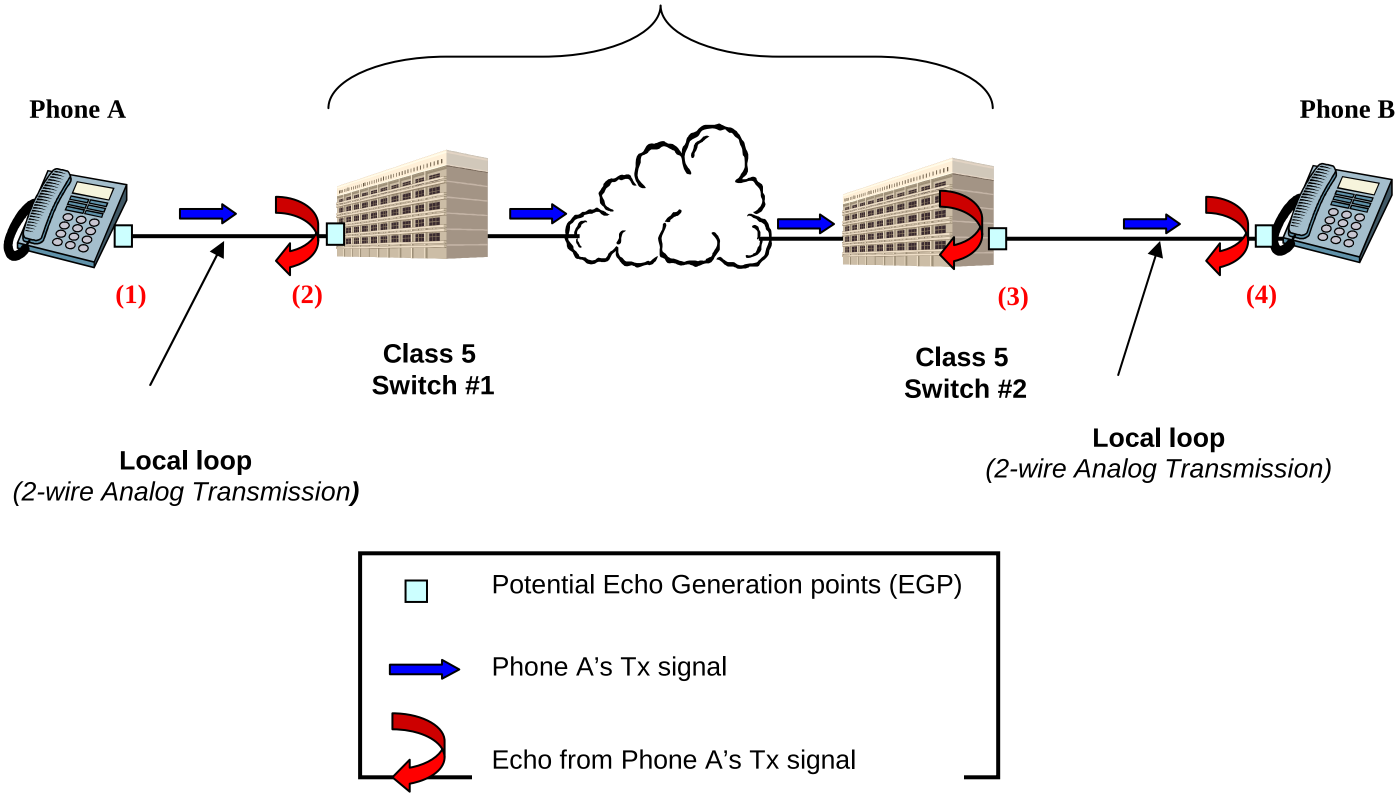


Fig. 5: Line Echo At The Local Loop

To get a better understanding of the sources of line echo, some background information on the PSTN is required. The telephone network consists of two basic sections: (1) the switching and transport core and (2) the local loop.

The switching and transport section is responsible for the transport and routing of calls, call services, billing, etc. Within this section, all voice and data signals are transmitted digitally, using separate paths for the Tx and Rx signals. This makes it much easier to transmit long distance signals, allowing for the use of repeaters, microwave transmission towers, etc. The local loop consists of the "last mile" of copper that connects the central office to the subscriber. For illustrative purposes consider a simplified PSTN diagram.

Fig. 6: PSTN Switching/Transport Network
(4-wire Digital Transmission)


Phone A
Phone B
(1)
(2)
(3)
(4)
Class 5
Switch #1
Class 5
Switch #2
Local loop
(2-wire Analog Transmission)
Local loop
(2-wire Analog Transmission)
Potential Echo Generation points (EGP)
Phone A's Tx signal
Echo from Phone A's Tx signal

If we trace the Tx signal of Phone A in the diagram, the light blue squares indicate potential echo generation points (EGPs) due to either transhybrid imbalance or impedance mismatches. In this example, we'll assume that these EGPs actually produce large enough echoes to be problematic. You can see from the diagram that there are 3 places where echo will be generated for Phone A (EGP 2, 3 and 4). The echo at EGP 2 is a result of impedance mismatch of the hybrid to the line and results in an echo being sent back to Phone A. This echo is not problematic as the round trip delay is not very long and will not be noticed by the end user. This echo is also masked by the sidetone. The problematic echoes are the echoes generated at EGP 3 and 4 since their round trip delay will be large (>25 ms). It is characteristic for the echoes generated at the "far end" of the network to cause the most problems. The echo at EGP 3 is due to transhybrid imbalance, while the echo at EGP 4 is due to the impedance mismatch of the phone to the line.

To help eliminate the unwanted echoes, telephone service providers will typically install echo cancellers in the PSTN. In order to eliminate the echoes on Phone A's connection, an echo canceller would need to be installed just before or in the #2 Class 5 switch. The echo canceller is inserted into the 4-wire section of the network. Its job is to model the echo characteristics of the local loop section containing EGP 3 and 4 and, along with knowledge of the Phone A's Tx signal, cancel the echo generated at EGP 3 and 4 prior to transmitting Phone B's Tx signal back to Phone A. In VoIP systems, the echo must be handled completely in the VoIP gateway for calls originating and terminating at VoIP CPE gateway. For calls that must be routed to the PSTN, the PSTN will continue to handle the echo cancellation function.

Standards

Although placing VoIP calls over the packet network does not require a direct connection to the PSTN, it does not preclude a system designer from meeting the requirements of the PSTN. A VoIP residential gateway designer must understand the telephony requirements that exist today and apply them to the FXS and FXO interfaces of their systems. There are four primary standards that are essential for understanding the behavior and requirements in the PSTN:

1. GR57 specifies the behavior of a Digital Loop Carrier system from the tip/ring pair at the central office to the tip/ring pair at the subscriber end of the system.
2. TA909 describes the behavior of Fiber-to-the-Curb systems from a digital interface at the central office (T1 for North America) to the tip/ring interface at the subscriber end of the system. Fiber-to-the-Curb systems are typically 16- to 24-channel systems.
3. GR303 discusses the behavior of "integrated" Digital Loop Carrier systems. Behavior is specified from the digital interface at the Central Office (T1) to the tip/ring interface at the subscriber end of the system. Systems of this type tend to be large with several hundred tip/ring pairs at the subscriber end.
4. GR1089 specifies the environmental standards for telephone infrastructure in North America. This includes specifying immunity to lightning, power cross, and EMC. It also specifies limits to unintended RF radiation (EMI).

Summary

As interest in VoIP service increases, phones specifically designed for voice calls over the packet network (i.e. IP phones) will eventually replace standard analog POTS phones. Until that time, however, designers of VoIP gateways must consider how their equipment will interface to these standard analog phones. The FXS and FXO circuits provide a means to this end.

About The Authors

Kim Devlin-Allen (kda@ti.com)
As a product manager in TI's VoIP group, Kim is responsible for the strategy for interfacing analog telephony solutions to TI's VoIP gateway products. Since joining TI in 1995, Kim has been a product manager for analog modem, central office line card and CPE gateway products. She earned her BS in industrial engineering from Iowa State University and has an MBA from Southern Methodist University.

Glenn Yancey (g-yancey@ti.com)
As a systems designer in TI's VoIP group, Glenn supports product definition and direction for communication processors. He earned his BS in Electrical Engineering from the University of Texas at Arlington.

Part Three: Voice Quality Assurance For VoIP Networks

by David Jarrett and Keith Buchanan,
Senior Broadband Applications and VoIP Gateway Product Manager,
Texas Instruments Incorporated

One reason that VoIP technology is becoming more widely deployed, aside from the decreasing cost-per-channel of enabling technologies such as DSPs, is the ability of VoIP systems to match the overall service quality offered by traditional circuit-switched voice networks. Many incumbent local exchange and long-distance service providers use VoIP technology in the backhaul portion of their networks without the end user being aware that VoIP is involved.

These traditional service providers use techniques to manage service quality developed over the last 100+ years for circuit-switched networks -- namely, careful network design, and tracking of customer and network trouble reports. Service providers use well-understood rules to characterize service level in terms of voice quality (based on loss, delay, and echo), and in difficulty in establishing a call. Networks are pre-engineered to offer a certain level of service while taking into account these factors. Then, a service provider’s main tool to assess service quality while the network is in operation is based on trouble reports from users, as well as general network equipment failure notification through Network Management systems.

Voice quality is, in reality, the end user's perception of quality. Network performance characteristics will impact voice quality (as discussed below). Metrics such as Mean Opinion Scores (MOS) measure the subjective perception of voice quality, and analysis tools can be used to derive these metrics.

However, as VoIP technology gets pushed closer to the edge of the network with IP phones (wired and wireless) and residential voice gateways, VoIP service providers have a much more difficult time assuring the voice quality for their subscribers for two important reasons:

1. Lack of control over the underlying transport network -- e.g. when proving voice service from a residential voice gateway attached to another provider’s residential broadband Cable Modem or DSL service
2. Use of transport technology that can vary in quality -- e.g. using WLAN media to transport VoIP, especially when the subscriber is moving

Fortunately the increasing processing power in these edge devices, which has enabled them to support high-quality VoIP service in the first place, will also enable them to directly measure and troubleshoot issues with customer service quality. A service provider can make in-service measurements of the voice quality their end-users experience, and can also use this information to separate problems with their VoIP equipment from those with the underlying transport, and therefore help them more effectively address issues whether they own the entire network or not.

VoIP Network Issues

Factors that impact voice quality in a VoIP network are fairly well understood. While most of these can be mitigated with careful network design, good quality assurance tools both in the VoIP endpoint equipment and the network itself can allow these issues to be addressed with the best balance of effectiveness and cost. The level of control over these factors will vary from network to network. This is highlighted by the differences between a well-managed enterprise network vs. an unmanaged network such as the Internet.

Network operational issues impact network performance and will create conditions that affect voice quality. These issues include:

- Outages/failures of network switches, routers, bridges
- Outages/failures of VoIP elements -- call servers, gateways
- Traffic management during peak periods and virus/DOS attacks.

Scaling to very large networks increases exposure and places more importance on effective planning and implementation. The following are several factors that must be considered when planning, designing and deploying VoIP networks.

Delay -- Caused by processing in the endpoint equipment (and in the network), the collection of voice samples to implement voice compression, and the collection of voice (compressed or uncompressed) into network packets. One-way delays of 400 ms or more will impact the ability to carry on a normal conversation (ITU-T Specification G.114). Delay can be mitigated with efficient VoIP gateway and network design (e.g. prioritizing voice packets to minimize switching and routing delays), but also by selecting the appropriate packet length to lower packetization delay.

Jitter -- Caused by the variation in delay characteristic of packet transport networks. This is best mitigated by adaptive jitter buffer management in the packet receive path, to effectively remove the jitter before the voice samples are played out to the listener.

Packet Loss -- Caused by packet buffer or processor overload in the network or the receive VoIP endpoint, or by packet bit errors. Best mitigated using packet loss concealment techniques as part of the voice compression algorithm to replay previously received voice and/or comfort noise samples until new information can be received.

Echo -- Caused by voice energy “bouncing” off the circuit at an analog PSTN interface (i.e. line to a telephone). Echo that is sufficiently attenuated and/or that is delayed by less than 15 ms will not be noticed. Echo between 15 ms to 35 ms will give the speech a “hollow” sound, while echo delayed more than 50 ms will be distinctly heard and should be cancelled (ITU-T Recommendation G.131). Echo is exacerbated by the additional delay caused by VoIP, typically in the range of 50 ms to 100 ms. Mitigation requires robust echo cancellation solutions in the gateways between VoIP and the PSTN.

Vocoder -- Voice quality is partially affected by the voice vocoder used. While the PSTN uses pulse code modulation (PCM - G.711), VoIP systems widely use low bit-rate vocoders such as G.729. The most commonly-used vocoders have acceptable MOS scores. Wide-band vocoders, such as G.722, can actually support voice quality on an all-IP voice network greater than that of a traditional circuit-switched voice networks.

Voice Activity Detection -- VAD is a popular extension to voice coding schemes that further reduce bandwidth by eliminating packets that contain silence. This sometimes affects call quality by clipping the beginning of a talk burst. This effect can be mitigated by careful tuning of the voice detection algorithm.

Other factors that may affect voice quality include:

- Signal loss and dropouts
- Background noise
- Signal attenuation/gain changes
- Level clipping
- Physical interface (e.g. analog vs. digital T1/E1)

Example Deployment-Related VoIP Quality Issues

New networking technologies and deployment models are causing additional challenges that impact the ability of VoIP service providers to guarantee the highest levels of service quality in their deployments. Two such examples are where the VoIP service provider does not control the underlying packet transport network, and use of packet networks with potentially high delay and loss, such as 802.11 (WLAN) technology.

Example 1:
A number of independent VoIP service providers are entering the market, offering consumer residential voice services at extremely low prices. These providers will provide a home gateway designed to be connected to a broadband internet connection (i.e. DSL or cable modem service), and will operate the infrastructure gateway equipment to connect subscribers to each other and to the PSTN.

These VoIP service providers are typically completely independent of the broadband access providers, so that the gateways will have no interworking with the transport network to allow support for end-to-end QoS. Indeed, since the transport network includes the Internet, there is no way to guarantee any level of packet jitter, loss, or delay. Therefore, more aggressive measures must be taken in the home and infrastructure gateways to mitigate possible degradation due to these effects. In addition, it is critical that these devices also provide robust measurement and troubleshooting tools to allow the service providers to know about, and hopefully localize, quality issues.

Example 2:
Significant progress has been made in the Wi-Fi Alliance and the IEEE 802.11 working group to add QoS-aware features to the WLAN MAC, such as access categories to handle the QoS requirements of voice (and streaming video) applications, and admission control policies to ensure WLAN channels are not oversubscribed. However, the fact remains that WLAN media will have relatively high loss and delay compared to wired Ethernet,

due to:

1. RF interference -- interference from other devices using the WLAN frequency bands (2.4 GHz for 802.11b and g), such as cordless phones and microwaves
2. Changes in the RF path -- e.g. due to moving objects reflecting RF energy, or motion from the end-station itself (e.g. because the user is walking or driving)

Again, it is necessary to have robust diagnostic solutions in the VoWLAN handset and in the overall network to identify voice problems per-call, to enable service providers and network operators to identify and most effectively address problems as they arise.

Voice-Quality Measurement Tools

The ability to capture and report events is critical for managing network performance. These tools must be extended to managing voice quality, allowing operations to identify and correct network problems that impact voice quality. In some cases the cause of the problem may not be determined in real time, requiring off-line analysis. Captured information can be reviewed to determine the root cause.

The oldest and most reliable voice-quality tool is the listening opinion tests where human listeners rate call quality in a controlled setting (ITU-T Specification P.800). Overall results are compiled to produce a mean opinion score (MOS), which is based on a panel of listeners ranking the quality of a series of call samples on a scale of 1 to 5 (“Bad” to “Excellent”, respectively). An aggregate score of 4 or more is considered toll (i.e. PSTN) quality. While this test has the disadvantage of being subjective, expensive, and time-consuming to produce, it is recognized as the most consistent measure of voice quality available.

The bulk of subsequent activity in voice quality measurement has been on producing algorithms and tools that can objectively measure voice quality -- i.e. based on direct mathematical calculation on sound samples, rather than listening tests. Such tests can be roughly classified as active (or intrusive) and passive (or non-intrusive). In general, active tests perform calculations on test or simulated calls and thus intrude on normal network usage (or are conducted in lab environments), while passive tests can perform calculations on active calls in live networks without any interruption of service. The following will explore the relevant tests in the categories further.

Active/Intrusive Tests

As described in various white papers [References 1, 2], a wide range of research into automated, objected voice quality testing led to the development of a number of algorithms based on perceptual modeling. The most widely used of these are:

- PSQM – ITU-T P.861 http://www.itu.int -- perceptual speech quality measurement; automated scoring system, design for circuit switched network

- PAMS http://www.psytechnics.com/downloads/pams/PAMS_white_paper.pdf -- perceptual analysis and measurement system; Intrusive speech quality assessment tool; end-to-end degradation analysis of injected signal
- PESQ- ITU–T P.862 http://www.pesq.org/ -- international standard for measuring end-to-end voice quality according to models of the human preception -- recent standard for assessing voice quality; leverages the best of PSQM and PAMS algorithms; supports voice encoding, jitter, packet loss, time-clipping and channel errors

From Ref. [2]: "Techniques are based on psycho-acoustic science, and use a common approach in which a sample of voice is input into a network, and the subsequent output is recorded. The output sample is then compared to the input sample to produce a score that represents how well (or poorly) the network reproduced at the output the original speech. The two key features of these techniques are that the input and output signal are both modeled in a "perceptual" domain first, and then the comparison determines audio-perceptual distances or disturbances as a human would perceive them. The objective of each technique is to produce scores, like MOS, that reliably predict the results of subjective tests."

Much analysis has been done on the relative merits of these (and other) techniques. Suffice it to say that while these algorithms have evolved over time to better model more situations that may arise in packet-based networks (e.g. packet loss, variable delay), PESQ was designed to combine the best aspects of the previous ones, and is recognized as providing the highest degree of correlation to subjective MOS testing.

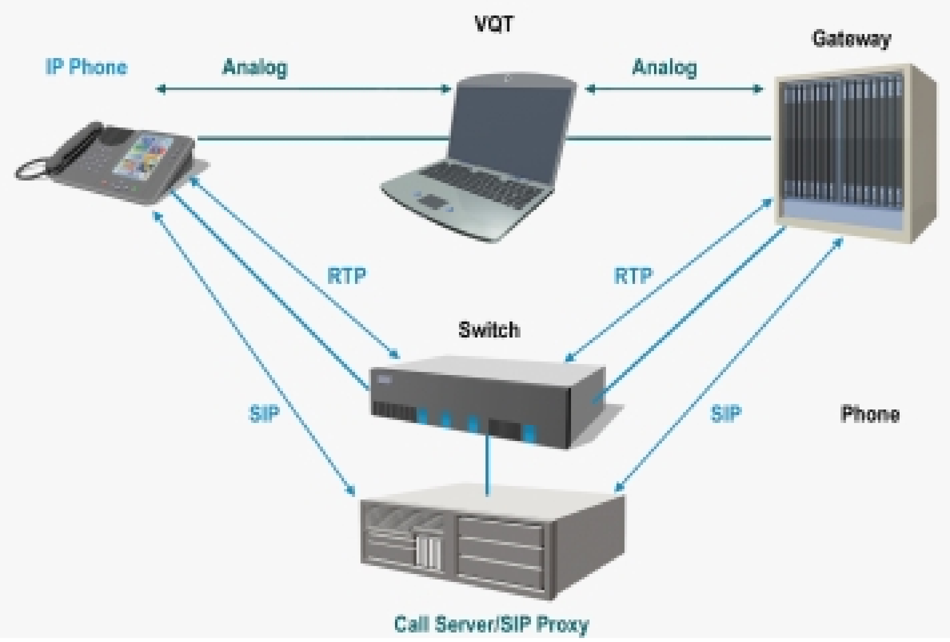


Fig. 1: Active/Intrusive Monitoring; Passive/Non-Intrusive Tests

Equipment from various manufacturers implementing these algorithms is widely used to test the quality of VoIP implementations at the component level and at the system level (Fig. 1). However, it is costly to use such equipment to measure the performance of active networks, since revenue-producing traffic must be interrupted to use it. In addition, while these algorithms can quantify deficiencies in speech quality, they do not produce information to help localize and identify the root causes of the situations causing the deficiency.

Passive tests, on the other hand, run in live networks without interrupting active calls and often use statistics gathered on active calls; they are therefore actually embedded into the VoIP equipment at the use site and in the VoIP service provider's network. As such, passive testing can therefore be used at lower cost as it eliminates both interruption to revenue-producing traffic and additional dedicated test equipment equipment.

Many tools used here are based on the E-model, as described in ITU-T Recommendations G.107 and G.108. The E-model is a transmission planning tool meant to account for a number of real-world factors to predict the performance of a network. The E-model calculates a transmission rating factor, R, calculated as:

$$R = Ro - Is - Id - Ie + A$$

where, *Ro* is SNR, including circuit and room noise;
Is is an impairment combination that occurs simultaneous to speech, including too-low send/receive loudness, non-optimal sidetone, and quantization distortion;
Id is a combination of impairments from delay, including talker echo, listener echo, and absolute delay;
Ie is an equipment impairment factor due to low bit-rate vocoders;
A is an advantage factor that accounts for the added convenience of different types of access. For example, mobile telephony has a higher advantage factor than wired telephony.

The ITU specs also describe how the R factor can be related to MOS.

Various vendors have adapted aspects of the E-model to support real-time calculation of call quality based on information about a call (e.g. jitter, packet loss, vocoder used). These calculations take minimal processing resources and can be combined into the overall VoIP DSP software load. In this way, they can perform the measurements on active calls through the gateways, on a per-channel basis (Fig. 2). Also, these measurements will be single-ended such that they don't depend on collecting end-to-end information about each call.

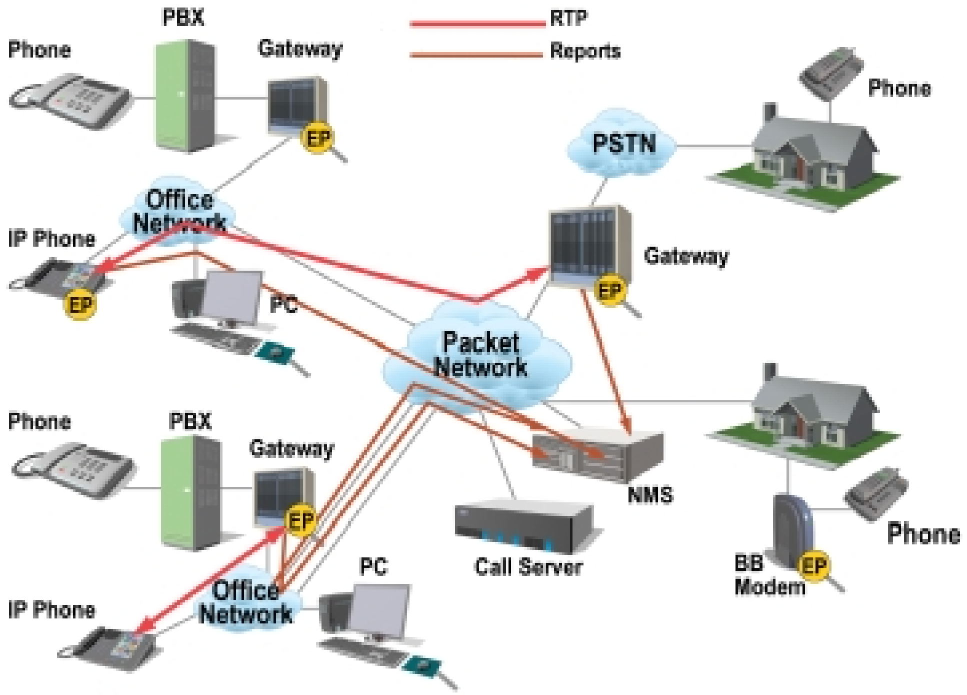


Fig. 2: Passive/Non-Intrusive Monitoring

While such uses of the E-model to perform passive measurements are not standardized, the ITU-T has recently standardized single-ended quality assessment measurements in ITU-T P.563.

Practical VQM Solution Aspects

Having passive measurements embedded in the VoIP equipment represents just the first piece of an overall solution for comprehensive voice quality monitoring. Once measurements are made in the VoIP gateway, this information must be reported to a Network Management system, where it can be used for problem detection and isolation. This and other information gathered from the network and the VoIP gateways can also be used for off-line analysis to diagnose the root cause of problems, so that they can be addressed in tweaks to the network or gateway configuration. The following will outline current work in these areas.

Reporting: RTCP XR

The Real Time Protocol (RTP; IETF RFC3550; http://www.ietf.org/rfc/rfc3550.txt) is the IETF standard for the transport of real-time data, including voice and video, in packet networks (including IP). It includes both a data part and a control part; the latter is Real Time Control Protocol (RTCP). RTCP offers general feedback of quality information in the context of a multicast group, as well as information to allow synchronization of multimedia streams. The Extended Report extension to RTCP (RFC3611;

http://www.ietf.org/rfc/rfc3611.txt) defines a format to transport information gathered in VoIP gateways in a standard and interoperable format.

Baseline RTCP includes sender and receiver reports that include some basic information (such as jitter, total packet count, and packets lost) about each call. RTCP XR defines a format to send report blocks that can be used for much more detailed information about RTP sessions. RFC3611 defines seven such report blocks; two are defined to carry summary metrics that are useful VoIP quality measurements:

1. Statistic Summary Report Block -- lost packets, duplicate packets, minimum, maximum, mean, and standard deviation jitter measurements, and packet TTL or Hop limit values by time interval (as defined by start and stop packet sequence number)
2. VoIP Metrics Block -- five categories of information including:
   - Packet loss/discard statistics -- loss rate, discard rate, burst metrics (burst density, gap density, burst duration, and gap duration)
   - Delay -- round-trip and end-system
   - Signal metrics -- signal level, noise level, residual echo return loss
   - Call quality metrics -- R Factor, listening quality MOS estimate, and conversational quality MOS estimate
   - Configuration parameters -- thresholds used, use of packet loss concealment, and use of adaptive jitter buffer

In addition, the specification defines a framework by which other implementation specific reports can be defined.

Having this information available, per VoIP endpoint, is clearly of great benefit in identifying potential trouble areas for entire networks, or even individual users.

Summary

Comprehensive QOS monitoring and management is required for VoIP services. Networks that are poorly implemented will adversely affect the end user experience and impact the broad acceptance of VoIP as a viable alternative to traditional telephony. Techniques and standards exist today for measuring and monitoring voice quality within the VoIP network elements. These tools must be included as the VoIP networks are deployed.

References

[1] Anderson, John, "Methods for Measuring Perceptual Speech Quality - White paper," August, 2002
[2] Anderson, John, "Addressing VoIP Speech Quality with Non-intrusive Measurements,"

About The Authors

David Jarrett provides field applications support for customer definition and design activities using TI's VoIP products. He earned his BS in Electrical Engineering from Kansas State University, and MS in Electrical Engineering from the University of Arizona. (djarrett@ti.com)

Keith Buchanan is product manager for TI's VoIP gateway product line. Keith has more than 20 years of experience in information technology and telecommunication and held system engineering and marketing positions with IBM and Ericsson before joining TI. He earned his engineering degree from Virginia Polytechnic Institute and State University. (kbuchanan@ti.com)

Part Four: VoIP Security Implementation

by Debbie Greenstree and Sophia Scoggins PhD
VoIP Business Unit
Texas Instruments Incorporated

Demand For VoIP Security

Customer Premises Equipment (CPE) -- IP phones and media gateways with VoIP capability -- is vulnerable to many Internet attacks, such as malformed frames or packet floods, both of which lead to Denial of Service attacks (DoS). Since DoS consumes significant equipment CPU processing cycles, this results in impaired voice quality in a real-time call processing scenario. VoIP CPE is also open for intrusion, monitoring, and alteration of the packet contents and destination addresses, and identity fraud in a non-managed environment. Therefore, VoIP security is a mission-critical element for the deployment of VoIP products. This article, the fourth in a series on CPE voice gateways addresses the implementation of security in such residential voice gateways.

Areas For VoIP Security

Fig. 1 shows a VoIP CPE gateway architecture consisting of two major components: Micro (voice application) and DSP. These can be inside an IP phone or in a separate box, such as a media gateway.

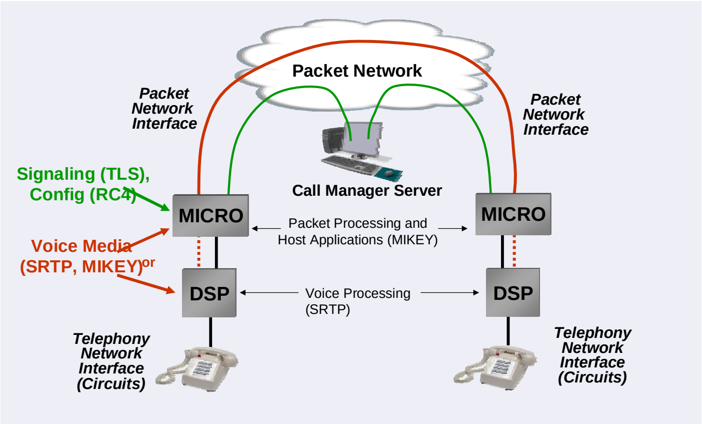

Packet Network
Packet Network Interface
Packet Network Interface
Signaling (TLS), Config (RC4)
Call Manager Server
MICRO
MICRO
Packet Processing and Host Applications (MIKEY)
Voice Media (SRTP, MIKEY)or
DSP
DSP
Voice Processing (SRTP)
Telephony Network Interface (Circuits)
Telephony Network Interface (Circuits)

Fig. 1: CPE Packet Telephony Security

The voice stream is packetized using IETF RFC1889 Real-Time Protocol (RTP), and as shown (using red dashes) is processed by the DSP using a voice encryption protocol and key exchange method. The encryption can be done by either the DSP or Micro. The key exchange method for voice encryption is between two Micros, relayed through the Call Manager/Server using IETF RFC 2327 Session Description Protocol (SDP).

The call processing signals (shown in green) are communicated between a Micro and a Call Manager/Server. In some situations, after a few messages between Micro and Call Manager/Server, the call processing messages may not go through the Call

Manager/Server any more, but directly between two Micros. The common call processing protocols are Session Initiation Protocol (SIP), H.323, and Media Gateway Control Protocol (MGCP).

In the architecture above, VoIP security can be divided into four areas: configuration, call control, voice streams, and data streams (see Fig. 2). Configuration is performed at the equipment startup stage with a configuration server. After configuration, the equipment may start data stream traffic. The data stream is independent to the call control or voice stream. When the equipment detects an off-hook signal, or incoming message, it starts the call control process with a Call Manager/Server. Once a call is established, the voice streams can be transmitted between two CPE gateways.

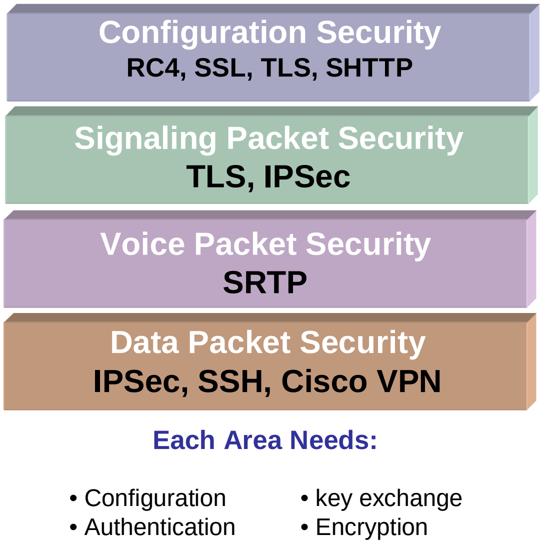


Fig. 2: VoIP Security Area

VoIP Security Components

Although the four areas have different security mechanisms, the basic security components are the same. The major security goals are authorization, authentication, integrity, privacy, and non-repudiation. In order to achieve these goals, the security mechanism often consists of configuration, authentication, key exchange, and encryption. Configuration is the initial stage to authorize the device in the network. Authentication may take place during configuration or at a later stage. Encryption is the mechanism for achieving integrity and privacy and requires a security key that can be statically assigned, or dynamically obtained, through key exchange. Non-repudiation can be achieved by a signature from the sender and/or sender and receiver reports, such as using the sender and receiver reports with the IETF RFC 1889 Real-Time Control Protocol (RTCP).

VoIP Security Performance Measurement

The major VoIP security performance measurement consists of the level of security, encryption delay, message delay, and processing power. Usually, the smaller the key size is, the less security, encryption delay, and processing power it has. A security key size less than 56 bit can be broken in three hours with sophisticated computers. 128 bit is the desirable security key size. A security key of size 192 bit consumes too much computation power. Although it does provide a high level of security, is not desirable for real-time call processing. The complexity of the security algorithm also impacts the level of security, encryption delay, and processing power. The message delay occurs during the authentication, key exchange, and call control process. In a real-time call processing application, delay can cause significant voice degradation and interfere with call establishment. Therefore, delays should be minimized. Any security mechanism introducing more than one second of delay is not suitable for real-time VoIP applications.

Encryption Protocols

The following summarizes common encryption protocols used in CPE applications, and their tradeoffs:

(Triple) Data Encryption Standard (DES/3DES)

The pioneers of voice encryption used IPSec with Data Encryption Standard (DES), Triple Data Encryption Standard (3DES), and Advanced Encryption Standard (AES). DES, 3DES, and AES are all endorsed by the US National Institute of Science and Technologies (NIST). DES uses a 56-bit key to encrypt blocks of 64-bit plain text. The key length is not long enough to provide security. 3DES uses 192-bit key. 3DES provides more security, but the computation time is too long so that it is not suitable for real-time voice processing.

Advanced Encryption Standard (AES)

AES uses a 128-bit key. AES provides much higher security level than DES, while the computational power is 3 to 10 times less than 3DES. AES is an ideal encryption protocol for voice and signaling systems.

Rivest Cipher (RC4)

RC4 was invented by Ronald Rivest at Rivest, Shamir, and Adelman (RSA). RC4 is a shared key stream cipher algorithm. The algorithm is used identically for encryption and decryption as the data stream is merged with the generated key sequence. The algorithm is serial as it requires successive exchanges of state entries based on the key sequence. Hence implementations can be very computationally intensive. RC4 is still the most common encryption method for encrypting configuration files.

Voice Encryption Protocol -- Secure RTP (SRTP)

SRTP is IETF RFC3711 [4]. SRTP provides a framework for encryption and message authentication of RTP and RTCP streams. SRTP adds two parts to the RTP header: authentication and encryption. Authentication is optional for SRTP, while required for SRTCP. Encryption is required for SRTP. Only AES encryption scheme is supported in SRTP.

Key Exchange Methods

The common key exchange methods are symmetric, public, hybrid, and Diffie-Hellman (DH).

Symmetric Key

This scheme uses only one key for encryption and decryption. Both ends of a phone call use the same key. The key can be generated by one end and distributed to the other end, or it can be assigned by a server to all parties in a domain. This method is not scalable, but it is the simplest method.

Public Key

This method uses two keys. The remote end's public key is used to encrypt the outgoing message. The private key is used to decrypt receiving message. This method is scalable, but needs 100 to 1000 times more computational power.

Hybrid Key

This method uses public key to encrypt the symmetric key. Once the symmetric key is received, it is used to decrypt the messages. This is the most efficient method and is used in many applications such as MS Outlook, Netscape Communicator, and secured data storage.

Diffie-Helman Keys (DH)

Two interacting endpoints must agree on a password in order for the call to go through. This is called the Diffe-Helman Scheme. One of the two CPE devices will pick a random number of base 2 and the other device has to match that number. There are five DH algorithms, or groups. The higher the group is, the more complex the algorithms are: which leads to a higher security level and more intensive computations. Due to the computation power required, the DH method is less used in voice applications.

IETF RFC 2401 Internet Security (IPSec) provides the security framework for key exchange, but refers to International Security Association (ISA) IETF RFC 2409 Internet Key Exchange (IKE) protocol for key exchange. IKE uses the DH key exchange method. IPSec has been very used in the pioneer voice applications.

An alternative to IPSec is to use the Multimedia Internet Keying (MIKEY) for key exchange for SRTP. MIKEY is currently an IETF draft, but is in the process of becoming a RFC. MIKEY requires the supporting of both public key and symmetric key methods, while Diffie-Helman (DH) is optional. Key exchange method will be carried in a SIP SDP attribute field. This field can be used for any key exchange method for media stream. MIKEY has limited implementations, but it is gaining industry attention.

Security Association (SA)

A Security Association (SA) is a virtual connection between two or more devices for the purpose of security. During the SA establishment stage, the devices perform authentication and exchange tokens or certificates, which are used to create encryption keys. Once the SA is established, some security mechanism will perform key exchange. In Fig. 1, there is at least one SA between each CPE and a Call Manager/Server. If there is a separate Configuration Server, then there will a SA between each CPE and the Configuration Server. There is also a SA between each pair of CPEs.

SA establishment is often time consuming, mainly due to exchanging messages. Therefore, SA establishment is recommended at the configuration stage between CPE and server. If the SA is expired, and requires renewal, it should be done when the devices are not in the call processing stage.

In addition to the SA between a CPE and a server, it is required to have the SA established between two or more CPEs. Pre-establishing a SA among all CPEs is not only unlikely, but also creates a meshed connection that will be difficult to manage in terms of memory and CPU processing power. Therefore, it is recommended to establish the SA among the CPEs on an as-needed basis. Since voice connections are often short, the SA can be terminated before it expires. It may be possible to reuse a previously established SA between two CPEs, if there is one. This can reduce some steps in the SA establishment stage.

VoIP Configuration Security

At start up, the customer premise equipment provides a pre-installed secure ID to the network configuration server. The configuration server responds with an authentication key. The CPE uses the authentication key to start the authentication process. Once the CPE gateway is authenticated, the configuration server provides an encryption key. From that point on the encryption key is used to encrypt all the messages between the CPE and the configuration server. The most common protocols used in this process are Rivest Cipher (RC4). Session Security Layer (SSL), Transport Layer Security (TLS), and Secure Hyper Tex Transfer Protocol (SHTTP).

SA establishment is part of the configuration process. Configuration is not unique to voice applications. However, while a data network may not require configuration at all, configuration for voice application is a must.

RC4 is a shared/symmetric key stream cipher algorithm. The key size is from 54 to 128 bit. The algorithm is serial as it requires successive exchanges of state entries based on the key sequence. Hence implementations can be very computationally intensive.

Security In VoIP Call Control Process

The VoIP call control or signaling system may use the authentication/encryption key generated at the configuration stage or use key exchange methods to obtain the encryption key.

Internet Security (IPSec)

The cable industry has been using IPSec using Kerberos key exchange method for the call control message in Media Gateway Control Protocol (MGCP). IPSec can be implemented under the IP stack and above the network driver, called bump-in-the-stack (BITS). An alternative is to implement IPSec away from the host, in the gateway, or router, or firewall. This method is called bump-in-the-wire (BITW). If IPSec is implemented in the IP stack, it can be used for all the applications in that device and the applications may not even notice it is in place. This is often implemented on a PC to set up a VPN to a Corporate Local Area Network (COLAN). If IPSec is implemented in a gateway, router or firewall, then many devices have to share the IPSec security association. This is often implemented between two office branches.

Transport Layer Security (TLS)

TLS 1.0 is derived from Session Security Layer (SSL) v3.0, but it is not backwards compatible with SSL. The SIP community first recommended using IPSec, but then changed to TLS. The old SIP spec was based on UDP, which required IPSec to provide more reliability, while the latest SIP specification is based on TCP. TCP provides sufficient reliability and therefore TLS over TCP does not cause reliability concerns. The TLS equips with key exchange function. Since TLS is above TCP, it often provides security association between two applications on two devices.

IPsec is used for long and reliable connection, while TLS is more for web-based applications with short and bursty traffic. With TLS, even after the SA is terminated, the application may reuse the previous SA information and re-establish a connection, which shortens the SA establishment time. IPSec does not allow reuse of the previous SA information to establish a new SA connection.

Security In Voice Processing

As stated earlier, the pioneers of voice encryption used IPSec with DES, 3DES, and AES. The latest standard voice encryption standard is the IETF RFC3711 Secure Real-Time Transport Protocol (SRTP) with AES. SRTP does not define what key exchange protocol to use. The industry trend is to use MIKEY for key exchange for SRTP.

Denial of Services (DoS)

DoS attacks are common in the Internet, and approaches to handling these attacks are not unique to VoIP. Highlighted below are some examples of DoS attacks and actions. There are public websites, such as CERT advisory board (http://www.cert.org/advisories/CA-1996-21.html or http://www.cert.org/advisories/CA-1997-28.html) which offer solutions to DoS attacks, as well as commercial products, which address this problem.

S.No	Attack Name	Scenario	Counter action
1	ICMP flood	High incoming rate of ICMP packets	Software restricts the number of packets to be received in time slot, if packet exceeds in defined time slot, log and drop the packets.
2	Teardrop	Not properly handling overlapping IP fragments.	Check IP fragments. Drop packets if they are not properly formatted.
3	Land	Source and destination IP address of packet is the same	RFC2267 -- Software input filter (for external traffic) does not allow packets through if the address is from internal. Software output filter does not allow packets through, if the source address is not from internal. Compare Source with destination IP address of packet, if same, log and drop the packet.
4	Ping to Death	High incoming rate of Ping packets	Restricts the number of ping packets to be received in time slot, if packet exceeds in defined time slot, log and drop the packets.
5	IP spoof, SYN flood	High rate of TCP SYN packets	RFC2267 -- Software input filter (for external traffic) does not allow packets through if the address is from internal. Software output filter does not allow packets through, if the source address is not from internal.

Open Issues

Although the industry has provided many solutions for VoIP security, there are still issues to be resolved. Most of the challenges come from managing the security keys. There is still no consensus on how to distribute the keys, update the keys, store the keys, and prevent them from being stolen. Meanwhile, the FCC has issued requirements for VoIP to comply with the Communications Assistance for Law Enforcement Act (CALEA). That means that VoIP service providers must provide a way for law enforcement agents to tap into the VoIP lines, or risk facing big fines.

Conclusion

Despite some of the challenges outlined, VoIP security is achievable now. With security in place, VoIP applications are expected to proliferate in the years to come.

About The Authors

Debbie Greenstreet is the Director of Product Management for CPE VoIP Gateways at Texas Instruments. She is responsible for product direction of CPE voice products, including VoCable, VoDSL and SME solutions. She has been working in the VoIP industry since its infancy, and has authored many articles and presented at numerous conferences on the subject. Ms. Greenstreet has over 20 years experience in the networking and telecommunications field in hardware and software design, as well as product management, at companies such as Hyundai and Raytheon. She holds a BSEE from the University of Virginia and has done graduate work in Computer Engineering at George Mason University. (dgreenstreet@ti.com)

Sophia Scoggins joined Texas Instruments in 2003 as a software system engineer. Earlier she had held many different positions (Director of Software Architecture, Sr. Product Manager, Architect, Sr. Software Engineer, and Research Assistant Professor) at companies, such as Nortel Networks, Siemens Efficient Networks, Coppercom, UMKC, etc. She holds a PhD in IE from TTU, is a PhD candidate and holds a MS degree in Telecommunications, Networking, and CS from UMKC, an MBA degree from ENMU, and a B Law from Taiwan. She holds two international patents and has published one textbook *Open Internetworking with OSI* and 45 conference, seminar, and journal papers. (sscoggins@ti.com)

References

For more information on Texas Instruments, and its VoIP solutions, visit http://www.ti.com/voip
Steve Burett & Stephen Paine, "RSA Security's Official Guide to Cryptography", McGraw-Hill, 2001.
Peter Thorsteinso, G. Ganesh, ".NET Security and Cryptography", Prentice-Hall, 2004.
"PacketCable Security Specificatio", PKT-SP-SEC-I11-040730, July 30, 2004.
J. Rosenberg & H. Schulzrinne, IETF RFC 3581- "An Extension to the Session Initiation Protocol (SIP) for symmetric Response Routing", Aug. 2003.
Mark Baugher, Ran Canetti, Lakshminath Dondeti, and Frederik Lindholm, IETF-DRAFT draft-ietf-msec-gkmarch-07.txt, "Group Key Management Architecture", Jan. 2003.
Kavita Jain & John Albert, IETF draft-jain-sipping-srtp-00.txt, "Using SRTP with SIP", Feb. 2004.
C. Jennings, IETF draft-jennings-sip-sec-flows-01.txt, "Example Call Flows Using SIP Security Mechanisms", Feb. 14, 2004.

J. Arkko, et. al, IETF draft-ietf-msec-mikey-08.txt, "MIKEY: Multimedia Internet KEYing", December, 2003.
T. Dierks & C. Allen, IETF RFC 2246, "The TLS Protocol version 1.0", Jan. 1999.
McGrew Baugher, et. al, IETF RFC 3711, "The Secure Real-Time Transport Protocol (SRTP)", March, 2004.
D. Maughan, et. al, IETF RFC 2408, "Internet Security Association and Key Management Protocol (ISAKMP)", Nov. 1998.
D. Harkins & D. Darrel, IETF RFC 2409, "Ineternet Key Exchange (IKE)", Nov. 1998.
http://www.webtorials.com/main/eduweb/security/tutorial/index.shtml
http://www.cert.org/advisories/CA-1996-21.html
http://www.cert.org/advisories/CA-1997-28.html

Part Five: Data Router Functionality

by Matt Harvill and Demetri Jobson
VoIP Business Unit
Texas Instruments Incorporated

This is the last of a five part series providing design considerations for the major portions of VoIP residential gateway products. A plethora of designs have emerged as offerings have moved beyond early Japanese markets and into mainstream US opportunities.

Growth in the residential gateway market can be attributed to the continued adoption of broadband in the home with almost 100 million users worldwide; the quality and functionality of VoIP continues to improve as technology advances in both hardware and software; also, the economics of migrating to VoIP continue to provide significant cost savings to both traditional telephony service providers and residential voice customers, while creating a path for new voice service providers, such as Vonage, to emerge.

As the residential VoIP market continues to grow, and mature, a variety of end product configurations have emerged, including solutions that not only integrate the voice functionality but also combine the home router as well. The early adoption of a terminal adapter-oriented product, providing only basic analog to IP conversion for VoIP service, is being replaced by this much more feature-rich voice gateway solution. Here we address the technical issues of such a design, as well as design and architecture tradeoffs and considerations, essential for an optimal end equipment configuration.

Data Router Basics

It is important to understand what end-product data router functions are required as they drive performance requirements, hardware/software specifications and affect the product architecture. Data routing directs data from the external WAN to a properly-addressed computer IP address on the internal LAN. Another function included in the residential router is a firewall to protect the LAN from corruption or sabotage through the WAN. The following are the typical features included in a residential data router:

NAT

Network Address Translation (NAT) is an Internet standard that enables IP addresses on private, internal networks (LANs) to be separate or hidden from corresponding public IP addresses. The NAT function provides the necessary address translations so that data can pass back and forth from the LAN to the WAN and vice versa, while protectively shielding the internal IP addresses from public view. The most common concept of NAT, used extensively in broadband routing, is an extension known as Network Address Port Translation (NAPT) and/or Port Address Translation (PAT). In the context of broadband routers, NAT usually implies the NAPT/PAT extension which introduces the highly desirable feature of one public IP address being mapped to many private IP addresses through the inclusion of a port identification number in addition to the LAN IP address. NAT (with NAPT/PAT implied) allows many additional IP addresses to be used internal to a LAN while appearing as a single public IP address.

There are a limited number of IP addresses available in IPv4, the current standard for IP addressing. IPv4 use a 32-bit address resulting in slightly over 4 billion unique addresses for the Internet; however, many of these addresses are reserved for special purposes leaving far less available for public consumption. Internet services providers usually allocate one address for each resident or entity and this single address was adequate when there was typically only a single computer in each household. As PCs have proliferated throughout the household, home networking has increased significantly in the form of routers and LANs. NAT has been the enabler, allowing multiple PCs or devices on a home network to appear as a single IP address to the public network, therefore consuming only a single address. This provides a layer of security by isolating the internal addresses from public access and view and also allows for internal addressing schemes and management without conflict to the public IP addressing model.

When a packet from the LAN is to be delivered to the WAN, the NAT:

- records the internal private LAN source IP number and source port number in its translation table
- replaces the source private IP address of the packet with its own public external IP address
- assigns a specific port number to the outgoing packet, enters that into the translation table, and replaces the source port number with this

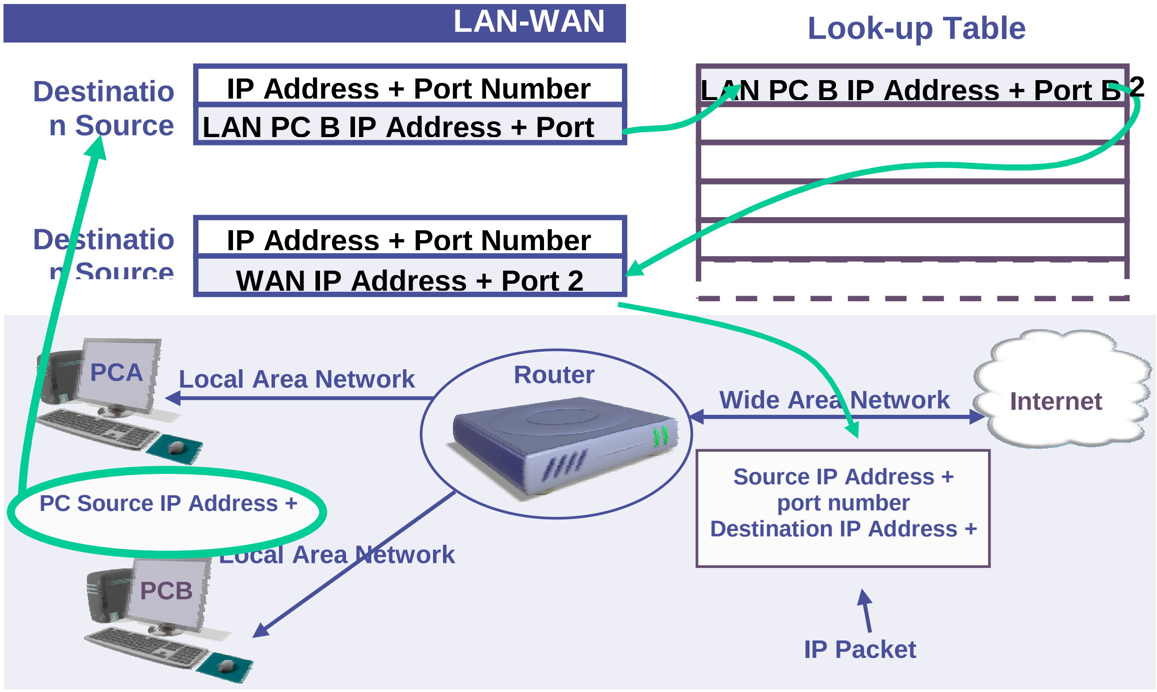


Fig. 1: Outgoing LAN - WAN Translation

When a packet from the WAN enters the LAN, NAT:

- checks the packet's destination port number
- finds out which inside machine was assigned this port number if it matches a previously-assigned source port number and discards the packet if a matching port number cannot be found
- rewrites, if a match is found, the destination port number and IP address with the original machines address and port number used for the packet on the inside that initiated the connection
- transmits this packet to the inside private network, for which it's intended
- maintains the translation table entries for as long as the connection is open

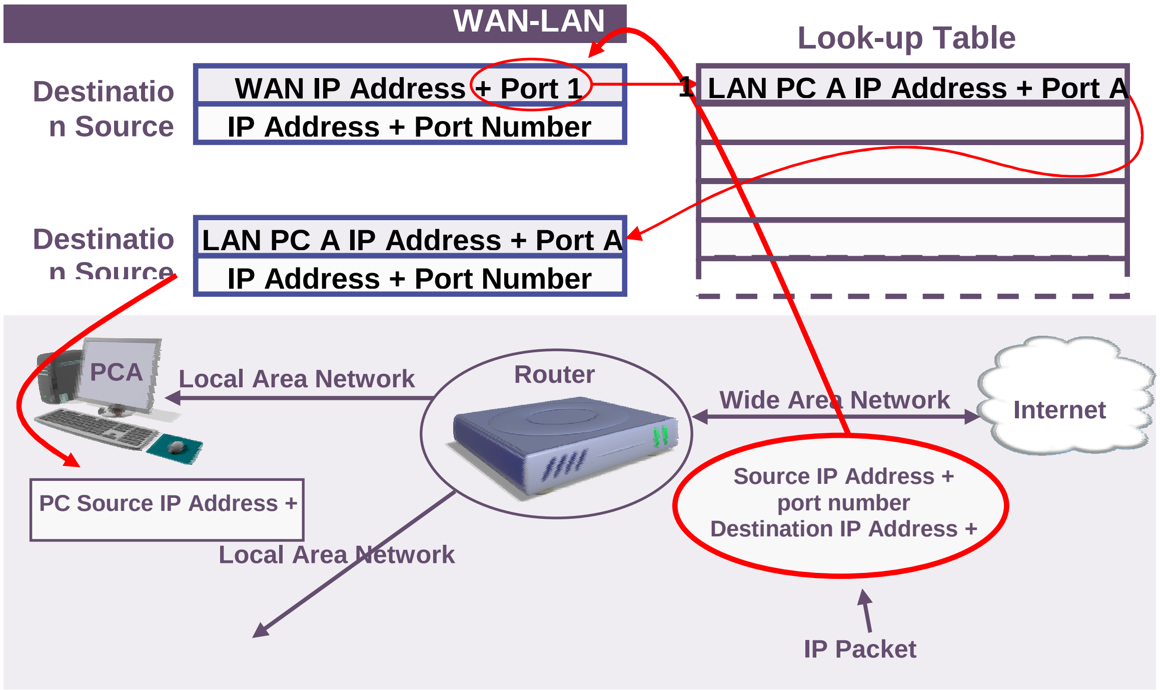


Fig. 2: Incoming LAN - WAN Translation

Firewall

A firewall is a software program or piece of hardware that enforces security between two networks determining which traffic to block and which to allow access. Firewalls are configured to protect against unauthenticated logins from the "outside" world and keep internal network segments secure.

Some of the features supported by a firewall include protection against:

- Remote login -- connection to a computer without approval
- SMTP session hijacking -- access to a list of e-mail addresses on a computer, sending unsolicited spam
- Operating system bugs -- remote access due to insufficient security controls/bugs
- Denial of service -- hacker inundating a server with unanswerable session requests, causing the server to crash

- E-mail bombs -- the sending of the same e-mail thousands of times until an e-mail system crashes
- Macros -- hacker creates macros that, depending on the application, can destroy data or crash a computer
- Viruses -- spreads quickly from one system to the next. Range from harmless messages to erasing all of a computer's data
- Spam -- links to web sites can accept a cookie that provides a backdoor to a computer
- Redirect bombs -- one of the ways that a denial of service attack is set up
- Source routing -- information appearing to come from a trusted source or even from inside the network

Depending on how vulnerable a network may be, a designer may want to enable all protection against external attacks. This maximum protection will take up extra CPU cycles and reduce performance of the data router. For more details on security issues, see Part Four of this series, above.

Router Performance Factors

For residential applications, the design of a voice gateway with router functionality is impacted by several performance factors, including:

- The number and type of data features supported
- The level of quality of service (QoS) supported for voice services

The router features will often impact the packet throughput performance. For example, if NAT is required, additional processing on the packets themselves is necessary to keep these NAT filtered packets at wire rate.

As this type of product combines residential data traffic with voice traffic it is important to maintain real-time throughput and QoS for the voice traffic (as detailed in the first article of the series). If the voice packets are not allowed to maintain a relative constant cadence, with minimal delay, due to high data traffic or large packet sizes between the LAN and WAN ports the voice quality will degrade, resulting in an unacceptable product architecture that will not muster market or service provider acceptance.

Design Process

Target Performance

The voice gateway designer must have a good idea of the requirements affecting performance, as stated in the previous section, before commencing with the product design. Whether you are taking an existing voice gateway product and adding data routing capability to the product, or initiating a voice/data router product from inception, it is essential to determine how much processing power is needed to meet requirements.

Another important factor in implementing an optimized voice/router product is cost. Residential voice gateways are now available in high-volume, mass-marketed retail channels and face aggressive BOM pricing. Finally, time-to-market is a critical consideration in this dynamic market segment. A single one month delay in product rollout can cost millions of dollars in lost opportunity with retail promotions that hinge on product availability for specific events, such as back-to-school and holiday promotions.

Voice gateways have different characteristics compared to data applications. Voice traffic is real-time data, which relies on timely reception of packets, ie, if voice data arrives too late, it might as well be lost, as it distorts what the listener hears. So, in a voice gateway, voice traffic must receive preferential treatment and have priority before data traffic. This priority comes in the form of QoS. With QoS implemented, the router processes voice or real-time-protocol (RTP) packets ahead of other data packets by examining packet type, and placing voice packets ahead of other packets in the transmit queue.

The combinations of voice and data in a gateway solution have different requirements regarding packet size and throughput. Voice involves a small packet size (typically less than 200 byte/packet, dependant on the codec) and requires a smaller portion of the bandwidth due to lower packet throughput requirements. For a voice payload of 10 ms the packet throughput is 200 packet/s in both directions. Data transfer usually involves much larger packets (1 – 1.5 kbyte/packet) with higher throughput and consuming more bandwidth. An example of a data packet could be an FTP download of an MP3 file.

Broadband rates in the residence will vary greatly depending on the service provided. In Japan, where there is a large presence of fiber-to-the-home (FTTH), bandwidth can approach 100 Mbit/s. In the US, where DSL and cable are typical, rates under 24 Mbit/s will be the maximum (with a practical bandwidth between 500 kbit/s and 5 Mbit/s). These varying broadband access rates are important data points in determining product performance and architecture.

Packet throughput is characterized in packets per second that ingress or egress from the LAN port of the gateway/router to the WAN port. The term “line rate” or “wire speed” routing implies that the router device can achieve effective maximum throughput or bit/s through the Ethernet port. For example, for a 100-Mbit/s Ethernet interface, the wire speed throughput would be 100 Mbit/s. It is important to understand what data packet size the intended gateway is expected to operate at, and at what associated packet throughput it will be expected to achieve. For example, if the data router is connected to a DSL modem in the U.S., where downstream rates are significantly less than 10 Mbit/s, designing a data router of 100 Mbit/s is excessive for the intended deployment.

With the desired routing performance at a given packet size understood, along with the voice requirements (including budget for performance and feature growth) the design engineer is now ready to consider various architectures for the residential voice gateway with router functionality.

Architectural Options

The first product architecture is one of discrete components collectively performing the voice gateway solution along with the data routing functionality (see Fig. 3),. This architecture allows the DSP to be sized for the voice performance, which is typically where the telephony media processing is performed. The RISC can be sized at a relatively high frequency to assure the appropriate megahertz are available for the routing performance. However, this approach does not typically afford a "one stop shopping" hardware/software solution; the designer must find the devices, voice processing software, and routing software from a variety of sources, leaving a significant amount of work developing the glue software and performing the integration to the design team. This will affect the time to market as well as overall solution robustness. The discrete solution, representative of early VoIP implementations, will typically bear a high total cost due to lack of solution integration.

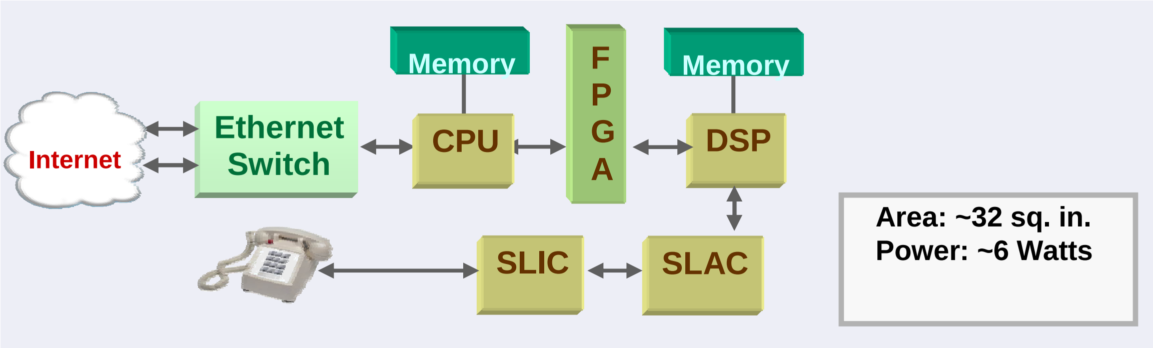


Fig. 3: Discrete Implementation Voice Gateway/Data Routing

The prevailing architecture in the residential gateway market is an integrated VoIP chipset (see Fig. 4).

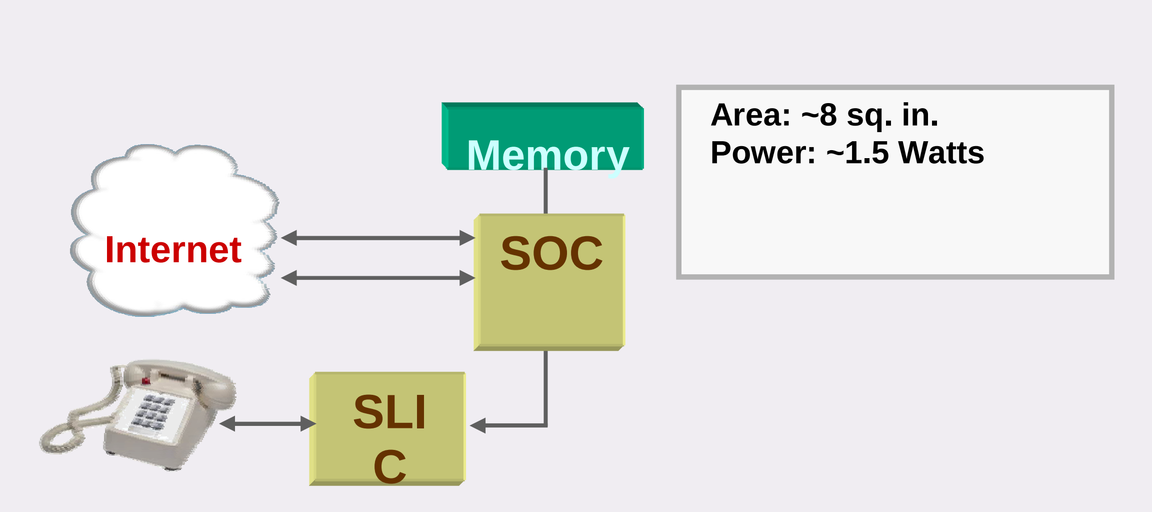


Fig. 4: Integrated VoIP Chipset

These devices typically include the DSP function for voice/telephony processing, the RISC processor for network and telephony protocol processing along with general device management, all in a single, integrated solution. The RISC processor will also be used to perform routing functionality. These integrated devices are available in a variety of speeds and performance levels; some are designed for basic voice gateway functionality, while others have the capacity to perform routing as well. This architecture is optimized for VoIP applications and often has the DSP and the RISC processor running at different frequencies, offering optimized performance for robust voice and routing.

Yet another VoIP architecture has emerged in some residential products. It consists of a higher-speed RISC processor that executes not only the telephony, network protocol and router functions, but also the voice processing operations (typically run on a DSP). This, of course, requires higher speed and, hence, more expensive processors. To run the voice function on the RISC, it typically takes twice the amount of RISC MHz than it does DSP MHz, so the designer must be careful in determining adequate MIPS sizing for the overall VoIP application as well as the router function, and to plan for functional growth.

Additional Design Factors

Board level costs/design complexities will be affected by the chosen architecture. Higher-speed processors may have higher-speed memory accesses and higher-density memory busses, subsequently requiring more challenging layout efforts. They may also be more subject to EMI issues. This could require additional shielding or even, possibly, a higher number of PCB layers (a usual goal is four layers). Such design complexities should be taken into consideration when comparing overall product costs and time to market.

Another design consideration is the number of LAN ports on the product. Currently most home routers are shipped with four-port LAN switches. In most households there are usually no more than two PCs, and with the acceptance of WLAN, there is little reason for additional LAN ports (although there is sometimes a need for at least one LAN port for configuration if the WLAN port needs setup).

Software image and execution size play a role in the overall bill of materials cost, in addition to maintenance and development costs. The FLASH and SDRAM components on residential gateways are a significant portion of the overall BOM. RISC-only based solutions, for example, typically have a larger FLASH and SDRAM consumption so this must be taken into consideration at the overall product level cost.

Conclusion

With the growth and interest in consumer VoIP, residential voice gateway designers are being challenged more than ever to come up with complete integrated voice/data solutions. Customers want robust voice quality and features, coupled with complete data functionality with routing, firewall and NAT services, to name a few. The overall development of such a solution can be greatly enhanced by understanding the relationship between these design requirements and the pressure to deliver a low-cost solution to the market. Fortunately the VoIP market has matured and evolved over the last few years with more tightly-integrated voice and data silicon solutions now available to meet designers' demanding requirements.

About The Authors

Matt Harvill is Product Manager for Texas Instruments' VoIP products. (mharvill@ti.com) Demetri Jobson is Systems Engineering Manager for Texas Instruments' VoIP products. (jobston@ti.com)

as published in...

analog ZONE