

# Designing intelligence into door-entry and security systems

DESIGNERS NOW HAVE GREATER OPPORTUNITY THAN EVER BEFORE TO DESIGN MORE INTELLIGENCE AND FLEXIBILITY INTO PERSONAL-SECURITY SYSTEMS BY USING MATURE AND PROVEN STANDARDS-BASED TECHNOLOGY. LEARN ABOUT SOME OF THE KEY CHALLENGES TO IMPLEMENTING HIGH-QUALITY, COST-EFFECTIVE, AND SECURE DOOR-ENTRY AND VIDEO-MONITORING SYSTEMS USING IP VOICE AND VIDEO TECHNOLOGY.

Personal security at home, on the street, or in a military setting is a permanent and growing concern; as a result, the products and technologies that address this security continue to see rapid market growth. A significant part of this market uses IP (Internet Protocol) video-surveillance systems—from extensive camera installations in metropolitan areas to small businesses and residences. Although this article refers to “door-entry systems,” keep in mind that you can apply this technology to many markets—video doorbells, video baby monitors, and even military applications. Door-entry and video-surveillance systems have become more realistic propositions at the low end of the market in both deployment and overall cost, mainly because of technological advances that support lower prices. These advances are largely due to both the widespread use of IP-based packet networks and the quality improvements in hardware and software.

Many of today’s door-entry and security systems employ analog, point-to-point connections feeding voice, video, or both over relatively short distances using expensive coaxial cabling. The costly addition of switching units to reroute the signals provides some flexibility, but overall flexibility is limited. Packet-based networks, however, provide a high degree of flexibility versus legacy analog systems. IP-based door-entry systems now enable viewers to observe images from any number of surveillance cameras that connect to their networks. They also now provide IP-based voice- and video-intercom capabilities both within and outside a LAN. The door-entry terminals that connect to the Internet can double as access points, so that service providers can use them for browsing or revenue-generating services. Today’s IP-based door-entry systems can also manually record video or record when certain events trigger the equipment and can route this recorded video

for storage at any point on the network. Some of these security systems also provide a “concierge” service that intercepts late-night calls and reroutes them to eliminate disturbance from unwanted visitors. Product design and development of these systems are also efficient because designers can reuse hardware and software for a range of applications and scenarios, such as in apartment complexes and industrial facilities (Figure 1).

The key drivers of this technology are the growth of broadband and the trend toward IP networks and IP-based communications. Also consider that you can use the same silicon devices in VOIP (voice over Internet Protocol) and video telephony in door-entry and surveillance systems. This extensibility provides a considerable advantage in reducing per-unit

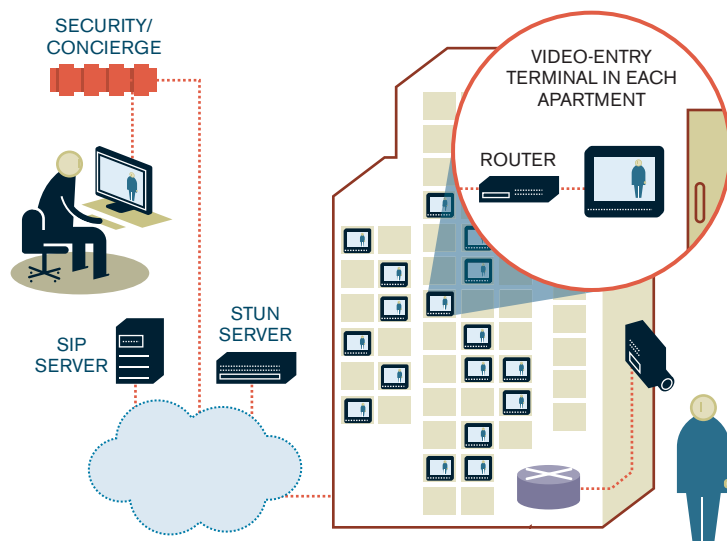


Figure 1 Some IP-based door-entry systems provide a “concierge” service that intercepts late-night calls and reroutes them to eliminate disturbance from unwanted visitors.

costs, as well as the benefit of a solid technological foundation for equipment manufacture and network reliability.

### ENABLING NEW APPLICATIONS

Perhaps most significant in the rapid growth of the IP-communications revolution and of IP-connected multimedia devices is the trend of embedding more and more “intelligence” at the edge of the telecommunications network. When VOIP services first emerged, media gateways resided in the core of the telecommunications network, which converted the PCM (pulse-code-modulated) voice from the PSTN (public switched-telephone network) into IP packets and vice versa. PBXs (private-branch exchanges) soon adopted the use of this technology in office or apartment buildings, handling comparatively fewer channels. Today, VOIP typically finds use as close to the edge of the network as is possible: at endpoints at which generation or reception of voice and video data occurs, such as within a telephone in the consumer’s home. Endpoint devices include cameras, handsets, monitoring stations, and home terminals. The components of the endpoint are the main focus of security-communications-equipment manufacturers. The telephones themselves can now connect to the IP network and make VOIP calls; with the increased bandwidth

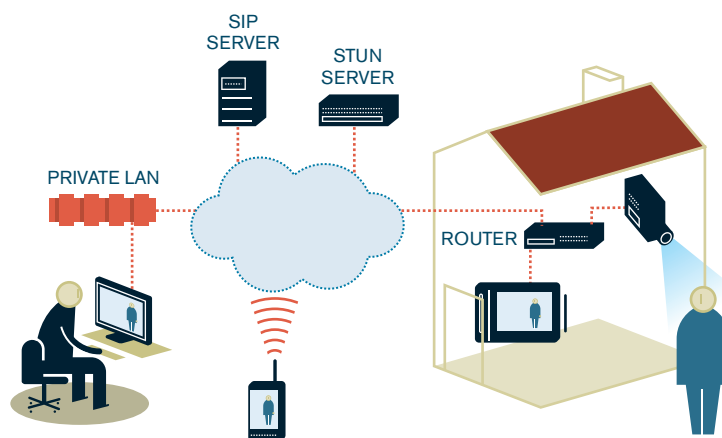


Figure 2 IP networks deliver flexibility to the end user.

that broadband lines offer, you can now make V2IP (voice-and-video-over-Internet Protocol) calls from these endpoints and use almost-identical endpoints as intercom units within video-doorbell and -security systems.

This migration to the network’s edge and into endpoint devices is largely a result of the industry’s acceptance of the SIP (session-initiation protocol) as the signaling and call-setup protocol of choice for IP-based communications. SIP’s ability to establish rich, multimedia communication sessions and enable features such as security and network intelligence—that is, “presence”—is lending itself to a host of other applications outside the traditional telecom function.

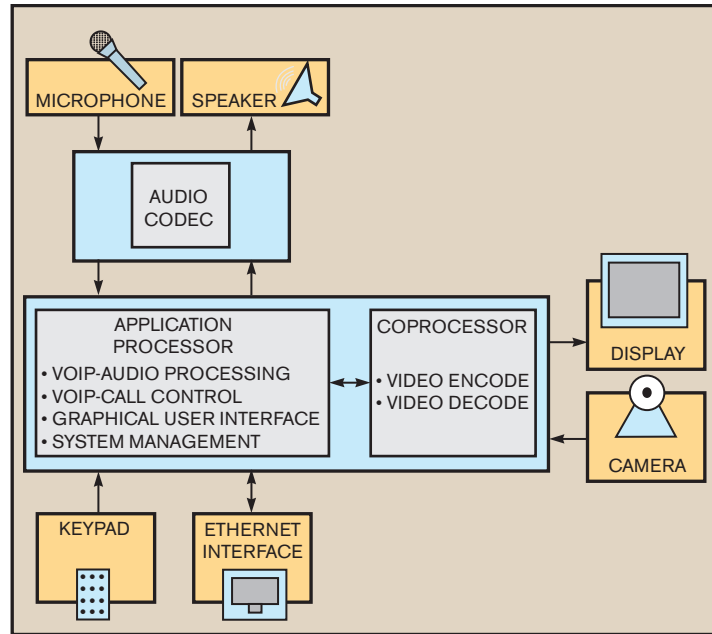
Figure 2 shows an IP network with the key elements for a variety of V2IP communications, including the endpoint; the SIP-registration server to identify and direct communications sessions between SIP-based devices across an IP network; the STUN (simple-traversal-of-user-datagram-protocol-through-network-address-translation) server, which helps the SIP signaling operate correctly even when the endpoints are behind a NAT (network-address-translation) unit; and the router for routing IP data around the network.

### CAN YOU SEE ME NOW?

The endpoint in a video-doorbell or -security system is typically wall- or desk-mounted in an apartment or an office. It comprises a microphone that connects to amplification and analog-to-digital-conversion circuitry, typically within a codec device; an amplified speaker that connects to digital-to-analog-conversion circuitry, typically within the same codec device; an analog or digital Web or surveillance camera; a video display, such as an LCD; an Ethernet or wireless-LAN interface; and a processing device, which includes a CPU with the processing capability for V2IP and any other applications the endpoint requires. A coprocessor within the device handles the video coding. The endpoint also includes V2IP software running on the CPU and in the application layer of the operating system. This software requires audio- and video-media processing and SIP-call control (Figure 3).

Most endpoints run some form of embedded operating sys-

tem on the CPU. The operating system can be proprietary, but the consumer-electronics and communication industries tend to use embedded Linux or Win CE. To use Linux or Win CE in an endpoint, a manufacturer creates or licenses a BSP (board-support package) for the CPU, which includes a root-file system with files and directories for applications, drivers, and configuration files; a kernel, which manages the CPU, memory I/O, and other resources; and driver programs for controlling peripherals. The operating system controls its peripherals, such as the audio codec, camera, and video display, through drivers. Linux provides standard interfaces to audio, including OSS (Open Sound System) and ALSA (Advanced Linux Sound Architecture), and video, including V4L2 (Video for Linux 2), to ease



**Figure 3** The endpoint in a video-doorbell or -security system comprises a microphone, an amplified speaker, an analog or digital Web or surveillance camera, a video display, an Ethernet or wireless-LAN interface, and a processing device.

portability between devices, and many manufacturers of peripherals provide free Linux drivers with their products.

Many silicon vendors provide reference platforms for their processing device, which customers can use to help accelerate their endpoint-hardware design. These platforms usually include most of the peripherals for the design of an endpoint, along with manufacturing information. They also typically include a BSP in source code, so that designers can use the reference platform as a starting point for software and hardware development.

Before a call can take place, you must be able to set up a connection

between two endpoints, such as two intercoms within an apartment building. For small networks, you can make connections

using the IP addresses alone or using a proprietary signaling protocol. However, for large networks, particularly when the two endpoints connect over the Internet, for example, it is better to use SIP because it intelligently handles the negotiations to securely and reliably establish these communications.

A SIP endpoint must first make itself known to a SIP-registration server, which is a means of discovering and identifying other SIP users in the network. The SIP user has an identification that uses the same format as an e-mail address—sipuser@sipservice.com, for example, in which “sipservice” is a SIP-network provider. If one SIP user wishes to call another SIP user, the first SIP user sends an invitation that indicates the IP address to which the second user should direct media. The invitation should also indicate the capabilities of the endpoint, such as which audio and video codecs it supports. The SIP server directs the invitation to the remote endpoint using a look-up service. The remote endpoint then responds directly to the initiating endpoint with its own capabilities and contact details. After

these negotiations are complete, the call commences with voice and video communication.

Apart from simple connection of two parties, SIP offers other enhanced services, such as call forwarding, call hold, voice mail, and instant messaging. A voice and video call, such as one that a video doorbell would set up, comprises one audio stream and one video stream. These streams are separate in the network and much of the software they use, but you must combine and synchronize them at the endpoint to ensure a good user experience. In both cases, the data streams pass through similar processing networks.

The encoding path from the microphone and the camera to the network includes the audio/video driver, which receives sampled speech and video from a buffer that a codec device feeds. After this process, AEC (acoustic-echo cancellation) occurs. AEC supports full-duplex, hands-free, speakerphone capability, which requires tuning to the device's acoustic properties. After AEC, video

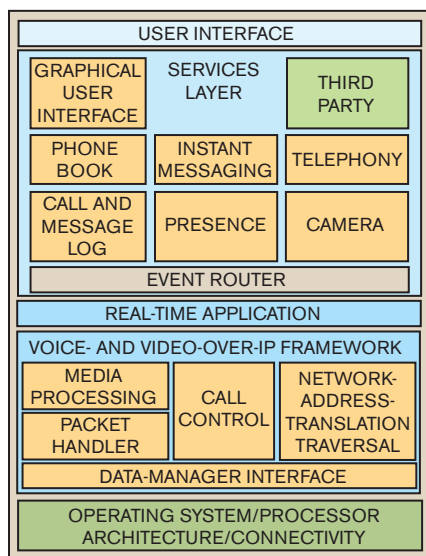


Figure 4 A framework that handles the lower levels of a VOIP and video-over-IP module consists of separate modules, each focusing on individual tasks.

preprocessing occurs. During this step, the video stream may require resizing, rotation, or mirroring. Next, a speech-

and video-coding process compresses the sampled data stream to reduce bandwidth requirements within the network. The bandwidth available is one factor in determining the codec and the video-image size. A common speech codec is G.729; H.264 is now in demand for video.

After this process, packetization segments the data stream into blocks for traversal of the network. Each of these requires an RTP (real-time-protocol) header to help with reconstruction of the stream at the remote end. The RTP header includes fields such as sequence number, time stamp, and SSRC (simple server-redundancy protocol), a number unique to the stream. UDP (user-datagram-protocol) and IP processing occurs next. In this step, the RTP packets pass to the IP stack and receive a UDP header corresponding to the destination address they arrived at during the signaling process.

After encoding, the decoding process takes place. Decoding operates on the data stream it received from the IP



network. First, during UDP and IP processing, received media packets pass through the device's UDP/IP stack. A jitter buffer assembles the received packets, using the RTP-header information, before decoding. The management of this buffer handles out-of-order and lost packets using PLC (packet-loss-concealment) techniques and is key to the device's QOS (quality of service). During audio and video decoding, the compressed data decodes to a raw format before passing on to the audio and video drivers.

## FUNDAMENTAL BUILDING BLOCKS

For the endpoint to function correctly in real time, these individual functions must take place in the correct sequence; all must coexist in the same processing device such that you can consider them as a single VOIP and video-over-IP module operating alongside any other applications the device requires. The V2IP module must be configurable in many ways, such as its management of drivers, to allow for low-power modes or for interaction with higher layers of the system, such as calls directly from an address book.

The best way to achieve these objectives is to provide a framework that will handle the lower levels of a VOIP and video-over-IP module. This framework consists of separate modules, each focusing on individual tasks (**Figure 4**). The objective of the framework is to speed the overall product design and development. A real-time application manages the modules in the framework; the application processes at the rate of incoming and outgoing data and determines the behavior of the modules by means of a set of APIs (application-programming interfaces). The use of APIs for interfacing to the un-

derlying framework allows you to abstract the application code from the scheduling of each of the components within each module, but it does have control over key components, such as codec, AEC, frame rate, image size, and bandwidth, to ensure the best user experience for the network conditions.

+ Go to [www.edn.com/ms4298](http://www.edn.com/ms4298) and click on Feedback Loop to post a comment on this article.

+ For more technical articles, go to [www.edn.com/features](http://www.edn.com/features).

The real-time application responds indirectly to user commands, such as making and answering calls. The user typically makes these commands using a GUI (graphical user interface) that itself has underlying service applications to support a phone book and its database or the logging of calls and messages, for example. The GUI and service applications behave differently from the real-time application in that they are event-driven rather than taking place in real time. For this reason, you must set up a communication path, such as the event router in the **figure**, between the two layers. The event router passes messages between the event-driven and real-time parts of the system. You must use a protocol and provide a queuing mechanism for these messages. You can use third-party-software vendors as sources for ready-made and tested software that provides all the applications from the GUI and services to the real-time application.

IP networks can overcome the connectivity disadvantages of point-to-point analog-door-entry and security systems. These networks can connect many endpoints and use the



same cabling for both video and audio data, and they can also control those endpoints. However, devices only recently emerged that allow sufficiently high-quality video to traverse these networks at a reasonable cost. These devices are now widely available due to the increasing popularity of VOIP and V2IP with consumers, and, because their functions are almost identical, they can find use within door-entry and security systems. Therefore, despite the fact that IP systems offer so much, their overall costs remain low.

The move from analog to digital packet-switched systems from a technical perspective is not trivial. However, you can gain much by outsourcing elements of the system and working with experienced software suppliers and silicon vendors offering devices targeting this market. A silicon supplier can provide reference designs to speed the hardware design and a BSP to afford a platform for application code. A software supplier can provide a framework with already-written code for the most specialized parts of the V2IP-software development.

Demand a software approach in which highly optimized speech and video codecs, echo cancellation, and call control all combine into a set of structured modules for use within a system developer's application. This step alone will result in shorter product-development cycles and the realization of significant time-to-market advantages.**EDN**

#### **AUTHOR'S BIOGRAPHY**



*Gordon Wilkinson has worked in the embedded-software and electronics industry for more than 18 years, starting with developing DSP-based instrumentation for nondestructive testing of materials. He then moved on to applications-engineering roles with LSI, Blue Wave Systems, and Motorola, specializing in DSP and VOIP. Wilkinson earned a doctorate in ultrasonic instrumentation and a bachelor's degree in electronics and computer science from Keele University (Keele, Staffordshire, UK). He is now a technical-account manager with Trinity Convergence, a leading VOIP- and V2IP-software developer. You can contact him at [gwilkinson@trinityconvergence.com](mailto:gwilkinson@trinityconvergence.com).*

