

# Do-It-Yourself VOIP

Gary Audin

## So you want to add voice to “this old network.” The enterprise IT handyman faces a host of tradoffs and choices.

Going to Home Depot to plan an extra room or remodel a home can be quite an experience. There are many choices, price ranges, competitive products and mutually-exclusive selections. You can put on an addition, finish the basement, or redesign the garage. Each choice will affect quality of life and also require compromises in order to keep the budget reasonable.

If an enterprise decides to implement VOIP, an analogous situation will occur. The products are not yet commodity technologies where all the vendors are virtually the same, as in 10BaseT Ethernet. Moreover, there are different ways to implement VOIP—as a software package, gateway, telephone- and/or router-based technology. One vendor’s product may be an efficient bandwidth consumer, while another’s can produce good-sounding voice during poor network performance. Does the desktop perform the VOIP function or do the network components? The answer can be “yes” to both.

The quality issue for VOIP is voice transmission. Is the voice clear and undistorted, with undetectable phone-to-phone delay? Do-it-yourself VOIP is, at least, the equivalent of remodeling your network but, more likely, making several additions tempered with compromises based on budget and quality goals.

There is no single solution for making a VOIP network perform better. You can throw money at the network and make everything bigger and faster. The network can be redesigned (the IP staff will like this one) to satisfy VOIP requirements. Careful selection of the products also can help. Finally, there are some potential problems that can be avoided.

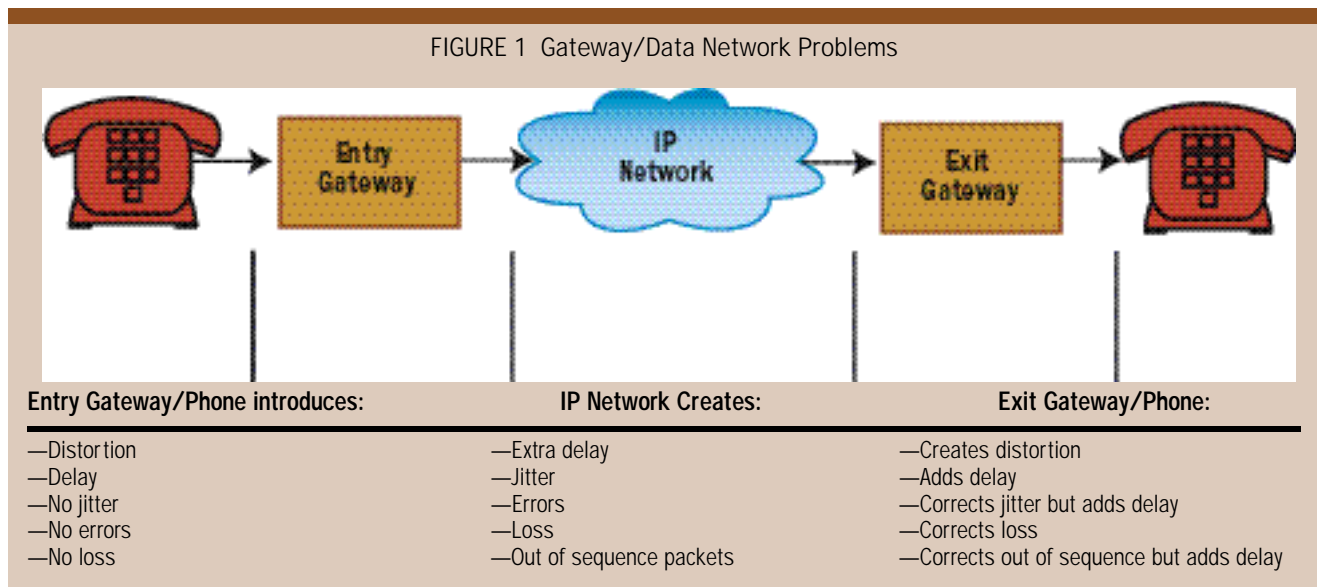
### Culprits In The Performance Equation

With the exception of cellular calls, we expect a network to deliver a consistent level of voice quality. But placing a voice call over an IP network involves using a network optimized for data, which introduces problems in three areas, discussed below, that do not exist on a voice network (Figure 1).

**n Entry devices—gateway, VOIP phone, etc.:** The entry gateway or VOIP phone introduces distortion when the voice is digitized and compressed using standards like G.711, G.729 and G.723.1. The lower the bit rate after digitization, the greater the distortion. Delay is introduced by

Gary Audin is president of network consulting firm Delphi Inc., and can be reached at [delphi-inc@att.net](mailto:delphi-inc@att.net). He is also the instructor for BCR’s “Internet Telephony” and “IP Convergent Networks” seminars ([www.bcr.com/seminars/default.asp](http://www.bcr.com/seminars/default.asp))

FIGURE 1 Gateway/Data Network Problems



Every problem in a VOIP transmission can be reduced. All it takes is money

the protocol processing and compression software, and this delay increases significantly as the compressed bit rate decreases. G.729 (8 kbps) adds a 12ms to 15ms delay, while G.723.1 (5.3 kbps) adds a 35-40ms delay before the compressed voice packet enters the IP network.

Jitter, or delay variance, can be introduced if a softphone program is used. The softphone software in a PC must compete with other resident programs for the PC's resources. To observe this conflict firsthand, try talking through a softphone when the PC is printing a PDF file.

If a gateway that is dedicated to VOIP is used, virtually no jitter will be produced. The gateway and VOIP phones do not introduce any errors. No loss will occur unless the bandwidth between the gateway and router is saturated, which is highly unlikely.

**n The IP network—routers and carrier transmission facilities:** Problems increase once the voice packet enters the IP network. Router processing takes 1-5ms per router, and traffic congestion can account for another 1-100ms per router, causing the IP network to produce considerable delay. Jitter occurs at each router because the voice packet must wait until the exiting router port can support the traffic. Each router can introduce 1-50ms of jitter. Jitter is also related to congestion, so increased congestion will equal increased jitter.

Transmission errors are almost always introduced by carrier transmission facilities, but the error rate is so low that it can be ignored for voice transmission. Packet loss is also caused by congestion. Greater congestion equals greater loss, and while most intranet designs limit packet loss to 5 percent or less, Internet loss rates can be as high as 30 percent.

Indeed, the most important factor in IP network performance is congestion management. IP networks are well designed and robust, and can reroute traffic around congestion points and network element failures. However, the rerouting does not mean the same performance will be delivered over the new path. In 1999, when four OC-192 cables of the Internet were accidentally cut, traffic was rerouted for some users from the East Coast to the West Coast through Sweden, producing a one-way delay in excess of two seconds. If the PSTN encountered such a cable cut there would be an increase in network busy signals, but the voice quality of the calls would not degrade. There are also far more backup facilities for the PSTN than for the IP networks when you view all the voice carriers combined.

**n The exiting devices:** The destination or exiting devices, gateway and VOIP phone also introduce problems when converting digitized voice packets back into analog voice. More distortion is created when decompressing the digitized voice back into an analog signal, and it increases if the vendors of the decompression software use different pro-

grams. This can be a problem even if the same standard is being used, and at last count, at least 11 different vendors were offering G.723.1 compression software. Not all of these different programs can interoperate correctly without adding distortion. The same delays are encountered in decompression as in compression of voice: 12ms to 15ms delay for G.729, and 35ms to 40ms for G.723.1.

Correcting the jitter involves using a jitter buffer to slow down the early arriving voice packets until the delayed packets catch up. Jitter buffers in some products are designed to compensate for as much as 200ms difference in packet arrival times. Packet loss requires further processing, which adds to the delay. A missing packet can be simulated by viewing the previous voice packets and creating a filler packet based on analysis of the previous packet combined with a knowledge of voice patterns. Alternately, the first packet following the lost packet can be used to simulate the missing packet. In either case, this is a skilled guess of the missing packet contents and adds a little more distortion on the voice call. If packets arrive out of sequence, further delay is incurred during the reorganizing process.

Not surprisingly, all of these problems can be reduced; all it takes is money. There also are many solutions which depend on the products chosen, bandwidth allocated, network redesign and the adoption of new router technologies.

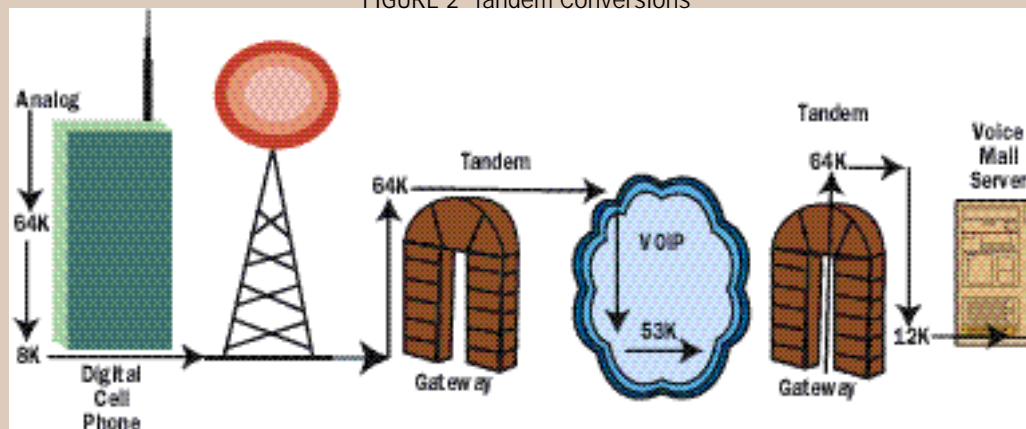
#### Hardware vs. Software Products

The conversion of analog voice to a digital PCM signal creates distortion, which is called quantizing noise, and it's not discernible to the human ear. The compression of the 64-kbps PCM stream to a lower bit rate causes reduced clarity and distortion as well as introducing delay. The greater the compression (5.3 kbps vs. 8 kbps), the longer the delay and the poorer the voice quality.

Generally, voice quality is better when the compression function is performed by hardware—a gateway or Etherphone—rather than a software-based PC softphone. That's because the compression delay is shorter with hardware, and if the PC is performing data functions and VOIP simultaneously, the conflict between the VOIP and data program for resources will cause voice quality to deteriorate and delay will be extended. Therefore, when making the switch to VOIP, it is better to avoid softphones and choose hardware-based gateway and phone equipment instead.

Another consideration when deciding between hardware and software is the choice of voice packet size—in bytes—and how much compression is performed by the product. A product that sends short packets of 20 bytes, rather than 48 bytes, means less delay and better compensation for packet loss at the receiving location. However, shorter packets increase overhead by 20-40 percent, because the protocol headers are a constant

FIGURE 2 Tandem Conversions



length and independent of the information (compressed voice) content size, therefore requiring more bandwidth per call. Alternately, using a compression algorithm of 8 kbps, as opposed to 5.3 kbps, can produce a voice transmission that will more easily tolerate performance problems. Processing delays at the gateway will be shorter and packet loss will be less noticeable. However, more bandwidth will be used for each call.

#### Buying And Managing Bandwidth

Allocating more WAN bandwidth always works because it reduces delay, jitter, packet loss and congestion. An advantage of installing higher-speed carrier circuits is that the cost per bit decreases as the bandwidth for a circuit is increased. A T1 bit costs about three times as much as a T3 bit, so the bill goes up, but not as fast as it would appear. The greater bandwidth should be applied to the trunks between routers. Routers are produced in various sizes, so the router you now own may not support higher-speed trunks. The routers may have to be replaced to support the higher-speed transmission.

LAN bandwidth is far cheaper to deliver for the enterprise. Most VOIP products operate over Ethernet at 10 or 100 Mbps. At 10 Mbps, 32-kbps per call and 30-percent maximum utilization of the LAN, the maximum number of calls is 93 at any time. This total assumes that no data is being transmitted on the LAN at the same time as voice calls. If the LAN bandwidth is to be limited to 10 Mbps, either a LAN switch should be installed or the LAN should be dedicated to voice-only operation, or both. Increasing the LAN speed to 100 Mbps will improve performance and increase capacity whether or not a LAN switch is used.

The expense of adding more bandwidth can be countered by reducing the bandwidth required. Bigger voice packets can significantly reduce the bandwidth required by 20–40 percent. Buying and configuring products that perform silence suppression and voice activity detection (VAD) can free up to 50 percent of the bandwidth, but this

assumes there really is silence on the phone call. Consider a call from an airport. The airport designers seem to locate the public address system just above the telephone locations, so there is never any silence to suppress.

Another technique that can reduce bandwidth consumption is header compression. The header for a voice packet contains a minimum of 12 bytes for the Real Time Protocol Header (RTP), and the User Datagram Protocol (UDP) header adds another 8 bytes. The IP header completes the overhead with a minimum of 20 bytes, for a total of 40 bytes of overhead without including any Layer 2 protocol (frame relay, ATM, Ethernet, etc.). The RTP Header may be longer when supporting multiple voice sources carried in one packet and/or for H.323 functions.

Header compression can be delivered in two forms: RTP only or RTP + UDP + IP. The RTP header can be reduced to as little as 2 bytes by reducing the sequence number, time stamp and synchronization source identifier fields. This header size reduction only affects the end points, not the routers. Compressing RTP, UDP and IP headers together can reduce the overhead to about 5 bytes depending on the actual header content.

This second technique is valuable, but it may require header decompression at each router so the IP header can be processed for proper packet forwarding. This adds to the delay going through the routers and increases the processing load of the router. The decompression at each router can be avoided by using MPLS (Multi-Protocol Label Switching) between routers. The MPLS label, not the IP header, is read for IP packet forwarding. The three headers (RTP, UDP, IP) remain compressed up to the last router or gateway.

Other than adding modest delay, header compression does not affect voice quality. The RTP + UDP + IP header reduces the bandwidth consumption for a Cisco gateway (AS 5300 with voice feature card) from 26 kbps before compression to 12 kbps after compression, assuming constant speech with no silence (see *BCR*, January

Aim for solutions that upgrade only the router

1999, “Voice over IP: Better and Better”). While a header-compression standard is currently being developed by the IETF, Cisco’s technique is proprietary, and as such will only interoperate with other products using their proprietary scheme.

Limit The Tandems

A tandem situation occurs when a voice call passes through multiple compression/ decompression stages (Figure 2). Each stage will noticeably add to the total end-to-end delay. When two digital cell phones are connected over the legacy network, the call passes through a tandem, and most callers can notice the increased delay. The compression/ decompression reduces voice clarity to the point where the distortion increases. This tandem situation occurs for several reasons.

There are four different digital cell phone techniques in the U.S. Other than Global System for Mobile Communications (GSM), it is unlikely that any of the VOIP standards or proprietary compression systems are compatible with the cellular digital compression.

Voice mail systems also use digitally compressed voice for storage. But there is no requirement for the voice-mail system to adhere to any particular compression standard, and no matter which is used, it is not likely to be compatible with the VOIP compression technique.

The compression techniques for VOIP gateways and softphones also vary, with many supporting proprietary compression techniques. At this time, the only reasonable way to perform the tandem function is to reconvert all compression back to the 64-kbps PCM signal before compressing with the next technique.

Tandems also will occur when the legacy 64-kbps PCM network is used for any of the following: traffic overflow from a VOIP network to the PSTN; connecting to a VOIP network at the remote location; backup to a failed VOIP component or failed router; or connection to legacy PBXs and phones.

The VOIP environment does not create the tandem situation. If all networks (legacy, VOIP, cellular) used the same compression/decompression technique, and were interoperable with each other,

and performed the voice compression technique once per call, the tandems would disappear.

Avoiding tandem situations may be impossible. Some PBX vendors, such as Avaya, suggest that a maximum of three tandems should be the goal. Although it seems unusual, an analog cell phone may sound better through a VOIP network than a digital cell phone, because it requires fewer tandems.

Remodeling The IP Network

The options available for renovating an IP network for voice include reconfiguring the routers and interconnecting circuits, upgrading the routers with new class of service (COS) and quality of service (QOS) functions, and modifying the gateways and VOIP phones and dividing the IP network into separate voice and data paths. (Table 1).

**Rearranging Network Components:** Routers are a bottleneck for voice traffic, as processing and congestion delays occur at each router. In IP terminology, one hop is passing through one router. The fewer the routers in a voice path, the less delay and congestion encountered, thus reducing the hop count. The “best” network would have only one hop, but this would reduce the circuit sharing that helps make IP networks more cost effective. Most intranet designers try to keep the maximum hop count to fewer than eight; fewer than five hops should be the goal. To ensure shorter delays with a large hop count, the circuit bandwidth utilization between routers should be less than usual, but this lower utilization raises the circuit costs.

A major change to the IP network would be the installation of ATM switches as the router-to-router backbone. ATM delivers short delay, supports COS and QOS and avoids IP processing by operating the OSI Layer 2. If a new IP network is being constructed or if ATM already exists within the enterprise, then ATM is a candidate for the backbone. Otherwise it is beyond what most enterprises need or can afford.

**Upgrading Router Appliances:** Any network performance upgrade, which is usually software, that can be restricted to the router will be more attractive than having to also modify the clients,

Technique	Standard	Proprietary	COS	QOS	Impact	
					Gateway Phone	Router
DiffServ	Yes	—	Yes	No	Yes	Yes
MPLS	Yes	—	Yes	Yes	None	Yes
RSVP	Yes	—	Yes	Yes	Yes	Yes
IPV6	Yes	—	Yes	Yes	Yes	Yes
L4Switching	—	Yes	Yes	Maybe	None	Yes
WFQ	—	Yes	Yes	No	None	Yes
ATM Backbone	Yes	—	Yes	Yes	None	Yes

servers, gateways and VOIP phones. Three possibilities exist under this condition: MultiProtocol Label Switching (MPLS), Weighted Fair Queuing (WFQ) and Layer 4 switching. See Table 1 for a comparison of techniques.

MPLS is a software addition to a router. A router capable of MPLS is referred to as a Label Switching Router (LSR). LSRs construct a virtual circuit that looks like a frame relay connection combined with COS and QOS. The addition of a switching label placed in front of the IP header reduces processing and delay at the router. This is effectively an OSI Layer 2 switching protocol tunneling through IP routers. The label includes the path identification, service description and a time-to-live field. The LSR is transparent to the end devices (gateways, VOIP phones), avoiding any changes to them. The LSRs, however, have to signal among themselves to set up the path and to remember the state (conditions setup) of the path. (For a more complete discussion see "MPLS: Dessert Topping or Floor Wax," in *BCR*, May 1999.)

Weighted Fair Queuing (WFQ) applies priorities, or "weights," to traffic types to classify the traffic into sessions, or "conversations." This is used to determine how much bandwidth is allowed for each session. Traffic flow is classified based on source and destination IP address, protocol (TCP, UDP, etc.) and port number (application identifier). WFQ is especially helpful for router-to-router trunks operating at T1/E1 speeds, as compared to First In First Out (FIFO) queuing. By this means, voice traffic can be given preferential treatment. Several variations of WFQ traffic management exist. No signaling is required for WFQ.

Layer 4 switching operates with similar traffic classification techniques as used in WFQ—IP address, port numbers and protocols. Layer 4 switches may also use RSVP filters, unicast and multicast forwarding and firewall processing. Some Layer 4 switches set up a form of label internal to the router to reduce processing delays and to make the packet forwarding decision faster and simpler. Another type of switch, a Layer 5 switch, inspects the content of the session to determine the packet forwarding policies. No signaling between routers is required.

**n Adding to Gateways and VOIP Phones:** The next set of alternatives will require changes to gateways and VOIP phones as well as router upgrading. Differentiated Services, or DiffServ, which appears to be gaining favor, produces fair treatment among flows but grants better treatment to some flows than to others.

DiffServ is a packet-forwarding technique used from one router to another based on delay, packet loss importance, cost and other factors. The specific treatment given a packet flow is based on the Per Hop Behavior, or PHB, as specified in the IP header and implemented by the router. DiffServ does not guarantee any quality of service; instead,

it is a class of service technique. It also does not increase the IP header; instead, it uses the existing Type of Service (TOS) field that is usually left empty. No signaling is required between routers for DiffServ; routers that do not support DiffServ software skip the field and treat the traffic as if DiffServ did not exist.

The Resource Reservation Protocol (RSVP) is a quality-of-service technique which assigns bandwidth to a specific packet flow. It is also a signaling protocol used to set up the reserved bandwidth path. It is set up in each direction separately. RSVP was standardized in 1995, but it has not been embraced as the final answer to QOS. It can be used in conjunction with MPLS and ATM networks. By itself, RSVP improves performance for specific packet flows. However, generally less than half of the available bandwidth can be reserved, so RSVP is not really fair to all traffic. (For more discussion on this, see "Lies, Damned Lies and RSVP," *BCR*, March 1997.)

Finally, we come to the next version of IP—IPv6. Though the IPv6 standards have been around for a while, there seems to be little interest in using IPv6's COS or QOS features. It involves too much change, will take too long to complete and there are many IPv4 enhancements that address the problems IPv6 was designed to solve. (Also see "Is IPv4 the Next Generation IP?" *BCR*, December 1998, and "Whatever Happened to IPv6?" in *BCR*, April 2001.)

In summary, techniques for redesigning the IP network are available, each with its own strengths, weaknesses and impact on network devices. Those most likely to succeed are DiffServ and MPLS, operating separately or together.

#### Measuring VOIP Performance


How good does the voice sound compared to the PSTN call quality? There are nine factors which contribute to a listener's satisfaction. Measuring voice conversation quality can be performed by people listening to calls, as defined in ITU Standard P.800, called Mean Opinion Score (MOS). Sound quality can be done using instrumentation and testing equipment as outlined in ITU Standard P.861, called Perceptual Speech Quality Measurement (PSQM).

The nine factors for voice quality are:

- n Distortion of speech.
- n Loudness (sound volume).
- n Background noise.
- n Voice loudness (volume) fading.
- n Crosstalk.
- n Network echo.
- n Echo-canceller performance.
- n End-to-end delay (phone to phone).
- n Silence suppression performance.

A good tutorial on these measurements can be found at <http://www.iec.org/tutorials>.

Mean opinion Score (MOS) is determined by a group of at least 30 judges who listen to sound



Nine factors  
shape voice  
quality



## Testing voice quality is not straightforward

bites, or “clips,” of prerecorded speech through various products and IP network conditions. The rating system used ranges from a score of 5, which is excellent quality, to 1, for bad quality. The goal is to get a rating of 4.0+. Standard G.711 has been rated by different testing groups as 4.5, 4.4, and 4.3. Were these better or poorer products for G.711? No, the O in MOS stands for “opinion,” not fact. If a rating is 3.84 vs. 3.75, is there really any difference? MOS scores can be this precise only when hundreds of judges participate in the same test. These figures should be rounded so that 3.81 is 3.8 and 3.75 is also 3.8. A human could not really tell the difference. This makes it difficult to compare results from two vendors using separate tests at different labs.

The Perceptual Speech Quality Measurement (PSQM) is the newer technique for measuring voice quality. It is becoming more popular with vendors in describing their performance. PSQM uses algorithms that automate sound quality evaluations. PSQM uses repeatable and objective calculations that include the subjectivity of the human factor for voice quality. Each of the speech representations has a weighting factor. For example, noise while speaking is perceived less by a listener than noise heard between the words spoken. The PSQM method can produce numbers that correlate to the MOS. The resulting score can be more precise than the human MOS scoring—the result of machine measurement as opposed to human measurement. A White Paper by Empirix discussing this subject is located at [www.empirix.com](http://www.empirix.com); click on “Resources.”

The difficulty in comparing different VOIP options is not only in the products themselves but also the nature and production of the voice quality test. Using a person who speaks fast but may also have an accent that is difficult to understand discussing a subject with an unfamiliar vocabulary through a marginal quality VOIP connection spells absolute dissatisfaction. The enterprise network personnel have several evaluation options. They can use real calls with a MOS testing environment under well controlled repeatable conditions, if the enterprise has the staff and testing conditions available. Alternately, they can simulate an IP network and its impairments for a MOS test of different VOIP products using an IP network simulator such as the one available from the National Institute for Standards and Technology (NIST), [www.antd.nist.gov/nistnet](http://www.antd.nist.gov/nistnet).

Another possibility is to simulate gateway/softphone VOIP traffic operation on an existing network to determine voice quality. Agilent, Empirix, and RAD, produce H.323 and SIP simulator products. A fourth alternative is to use real calls with VOIP monitoring devices and software to determine the voice quality. These products are also produced by Agilent, Empirix, and RAD.

Fifth, independent testing reports from companies such as Mier Communications ([www.mier.com](http://www.mier.com))

provide an unbiased repeatable set of product tests with evaluation scores. Performing your own MOS or PSQM testing is attractive, but it is not within the capability of most enterprises.

An inexpensive method for determining the IP network characteristics before VOIP is implemented can be produced using the IP ping function. A ping is a short packet transmitted by one IP device to another, for example, from a PC to a server. The receiving device loops (returns) the packet back to the sender. Ping is normally used to determine the operability of the remote IP device. If a ping is transmitted once per second for an entire week, calculations can determine delay, jitter, packet loss, and out-of-sequence packet reception. The transmitting device can measure the round-trip delay and determine the average, minimum, and maximum delays as well as jitter. Unreturned or out-of-sequence packets can be measured using this technique. The raw data can be analyzed and displayed in a spread sheet. An added function of ping is a feature called trace, which tracks the IP addresses of each device the ping packet passes through. If the trace shows a very consistent physical path, then the test results should be very accurate. However, if the trace varies considerably, then the calculated results will be less precise. The confidence level of the measurements will be lower.

Additionally, the IP network should be tested to determine its busy hour for data traffic. This is when IP network performance causes the greatest voice quality degradation due to congestion. Compare the IP network busy hour to the present voice network busy hour. If they overlap, VOIP traffic will be heaviest during the worst possible period for IP network performance. Performance will obviously be better if the IP and voice network busy hours occur at different times.

### Conclusion

Take a look at the book section at Home Depot. Some books are for beginners, some for experienced homeowners and a few for professionals. The library for VOIP varies considerably from basic to advanced, with the technical books written for VOIP developers. The “Dummies” book for enterprise VOIP has yet to be found. A Voice Over Data user group ([chair@voiceoverdata.org](mailto:chair@voiceoverdata.org)) has been created to exchange information and experiences and to help reduce the difficulties of proceeding with conversion to VOIP□

### Companies Mentioned In This Article

- Agilent ([www.agilent.com](http://www.agilent.com))
- Avaya ([www.avaya.com](http://www.avaya.com))
- Cisco ([www.cisco.com](http://www.cisco.com))
- Empirix ([www.empirix.com](http://www.empirix.com))
- National Institute for Standards and Technology ([www.antd.nist.gov/nistnet](http://www.antd.nist.gov/nistnet))
- RAD ([www.radcom-inc.com](http://www.radcom-inc.com))