Doing a VoIP Assessment with Chariot VoIP Assessor

by Susan Pearsall and John Q. Walker, NetIQ Corporation
susan.pearsall@netiq.com
johnq@netiq.com

Deploying Voice over IP (VoIP) successfully in an enterprise data network may have some unexpected pitfalls. In previous papers, we’ve examined key planning and design tips to help you avoid some of the frustrations associated with a failed or flawed VoIP implementation [1, 2]. This paper describes how to determine whether a data network is ready to run VoIP traffic well.

Chariot VoIP Assessor™, by NetIQ Corporation, is part of a suite of products designed to test, evaluate, monitor, and manage multi-protocol and multi-platform data networks. VoIP Assessor evaluates the network performance metrics that influence the quality of VoIP calls on a data network.
Even when they’re carried on the same network, voice traffic and data traffic can’t be handled the same way.

- They have different packet sizes.
- They are sent at different rates.
- They are buffered and delivered to the application differently.
- They must fulfill very different user expectations.

Although an e-mail message or a file transfer can be delayed by half an hour without exciting anyone’s notice, delays of a few hundred milliseconds can ruin a VoIP telephone call. And when you start to run VoIP across any given enterprise network, delays caused by other applications, overloaded routers, or faulty switches may be inevitable. Chariot VoIP Assessor is designed to:

- Assess VoIP readiness, determining whether an existing data network is ready to deliver quality VoIP calls in its current configuration.
- Evaluate VoIP call quality over the course of a few days, running hundreds or even thousands of simulated calls over the network and taking measurements.

VoIP Assessor sends simulated VoIP traffic between pre-selected points on a network for a period of time you select. While the simulated calls are running, VoIP Assessor takes measurements and calculates call quality scores. VoIP Assessor reports quantify the data collected over the course of the assessment to help determine a network’s readiness and capacity for handling real VoIP traffic.

**Why Run a VoIP Readiness Assessment?**

If you’re thinking about implementing VoIP or expanding the use of VoIP applications, you’re probably uncertain whether an existing network is ready to carry high-quality voice transmissions.

A VoIP Readiness Assessment is designed to systematically generate traffic loads that imitate a VoIP system’s traffic across the network. Such measurements provide information that can’t be gleaned from a pilot that simply uses an IP PBX and a few dozen IP phones [3].

You can use pre-configured defaults for system parameters, or you can tune them to see how various technical choices affect call quality and bandwidth consumption. For example, you can examine the effects of a half-dozen codecs representing various compression algorithms; you can also tinker with jitter buffers, delay between datagrams, and silence suppression.

The assessment software measures delay, jitter, and lost data, and produces a report showing call quality by day of week, location, network cause, and so on. You end up being able to tell what technical factors are affecting voice quality. The key thing is to get all these answers before you’ve spent a lot of money, time, and energy on VoIP deployment. You can work through all the data network issues so that by the time you actually start running the real VoIP piece of it, you have a data network that’s going to work well. You can also make cost-effective decisions about network infrastructure and application traffic once you know how voice over IP is performing.

Organizations typically deploy VoIP to lower telecommunications costs, by switching voice traffic to underutilized WAN and LAN links. However, it’s unlikely that any arbitrary data network is configured to handle VoIP traffic well. Don Peterson, President and CEO of Avaya Corp., recently observed that 85% of today’s router-based networks are not ready to support high-quality VoIP traffic [4]. Voice traffic is uniquely time-sensitive. It can’t be queued, and if datagrams are lost, the conversation will be choppy or even incomprehensible. NAT-enabled firewalls, slow or congested links, or improperly implemented QoS schemes are just a few of the factors that could inhibit or even prevent VoIP traffic from crossing a data network in an acceptable form.
**How Chariot VoIP Assessor Works**

VoIP Assessor determines how well voice over IP will sound on a network by assessing the quality of simulated VoIP calls. To assess call quality, VoIP Assessor sends realistic VoIP traffic across the network and measures the resulting flows.

VoIP Assessor relies on NetIQ Performance Endpoints, agents that are installed on every computer you plan to include in assessments. You deploy a Performance Endpoint or two at every site where there would be VoIP traffic. A small location would get one agent to measure WAN links; a larger site might have two agents: one at the edge to measure the WAN link, and another across the campus or large building, to measure traffic internal to the site. Although you can distribute endpoints anywhere on a network, including at remote sites, you don’t have to visit those sites to find out how well VoIP is performing there.

You set up an assessment at the VoIP Assessor console by first selecting the type of VoIP traffic to send. The key parameter to identify is the codec, which determines how an analog voice signal is digitized and the rate at which it is transmitted. You create call groups that act as senders and recipients of calls. A call group consists of two endpoints connected by a VoIP connector, which defines the type and number of calls to be sent between the endpoints on a specified schedule. The following drawing illustrates how VoIP Assessor and the endpoints work:

![Diagram of VoIP Assessor and endpoints](image)

In this drawing, endpoints that belong to call groups are designated by a telephone receiver symbol. All assessment parameters, including the codec to be emulated and the schedule of calls to be made, are saved in a database. This database also contains any results from the assessment after its run.

When you run an assessment, the VoIP Assessor console contacts all the endpoints in each call group. The console sends the endpoints a call script telling them what codec to emulate and a schedule to use (in the drawing above, dashed lines indicate the setup flows). The endpoints then send the information to their partner endpoints within each call group.

As the assessment runs (solid lines in the drawing), the endpoints take measurements and periodically send back results to the console, which stores them in the assessment database (dashed lines, with arrows pointing back to the console, indicate reporting flows).

The endpoints return results using the connection-oriented TCP protocol so that results aren’t lost. The simulated VoIP traffic uses the Real-time Transport Protocol (RTP).

**NetIQ Performance Endpoints**

VoIP Assessor runs with Performance Endpoints version 4.3 and higher for the following operating systems:

- Windows 98, Me, NT 4.0, 2000, and XP
- Linux for x86
- Sun Solaris (x86 and SPARC)

We don’t recommend using Windows 98, Windows Me, and pre-ULTRA Sun operating systems. The high-precision clocks on these systems are not sufficiently accurate for good delay measurements.
The Performance Endpoint software is free; anyone on the network can quickly install endpoints on their computers by visiting www.netiq.com/download/endpoints and downloading the endpoint software appropriate for the operating system they are using. At the same site, they can also download documentation explaining how to install and deploy endpoints for all the operating systems supported.

You can also download the endpoints and make them available on a shared network drive. Or you can distribute them via SMS, Microsoft’s Systems Management Server.

A single endpoint can participate in about 100 simultaneous VoIP calls without sacrificing call quality. (We tested using an endpoint running with Windows 2000 SP1 on an 800 MHz Pentium III with 256 MB of RAM.)

Score Calculation

VoIP Assessor calculates call quality based on a set of factors known to affect the perceived quality of voice over IP transmissions. A subjective factor is necessarily part of evaluating VoIP because a listener must be able to understand the received transmission, and both talkers must be able to tolerate the amount of delay between speaking and being heard (called “the walkie-talkie effect”), lost or fractured syllables, and echo that often impede the conversation.

That’s why the chief unit of measurement for call quality in VoIP Assessor results is an estimated Mean Opinion Score, or MOS. The E-model, ITU Standard G.107 [5], quantifies what is essentially a subjective judgment: a user’s opinion of the perceived quality of a voice transmission. After much study, the ITU determined which impairment factors produced the strongest user perceptions of lower quality. The E-model thus includes factors for equipment and impairments and takes into account typical users’ perceptions of voice transmissions.

VoIP Assessor uses the components of the E-model along with the following factors to calculate the MOS:

**Delay**

Delayed datagrams are perhaps the single greatest hindrance to VoIP call quality. This value includes all causes of delay between the endpoints.

**Jitter**

Jitter, or variations in datagram inter-arrival times, affects call quality. Excessive jitter reduces call clarity.

**Lost Data**

When a datagram is lost, you can lose an entire syllable, and the more datagrams that are lost, the more the clarity suffers.

VoIP Assessor reports also show how many calls could not be completed (“Unavailable” calls) and give you the objective measurements of delay, jitter, and lost data so you can independently judge quality for yourself.

By default, VoIP Assessor maps MOS estimates to three quality categories. You can change the way the categories are applied to conform to your own quality standards. A MOS of 5 is excellent; a MOS of 1 is unacceptably bad. The following table summarizes the relation between the MOS and user satisfaction:

<table>
<thead>
<tr>
<th>MOS Range</th>
<th>Quality</th>
<th>Meaning</th>
<th>Color</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.03 to 5.0</td>
<td>Good</td>
<td>Most or nearly all users satisfied</td>
<td>Green</td>
</tr>
<tr>
<td>3.60 to 4.02</td>
<td>Acceptable</td>
<td>Some users satisfied</td>
<td>Yellow</td>
</tr>
<tr>
<td>below 3.60</td>
<td>Poor</td>
<td>Most or nearly all users dissatisfied</td>
<td>Red</td>
</tr>
</tbody>
</table>

Here’s an example of how the overall call quality of an assessment is shown, using these default values:
VoIP Configuration

Before you start running assessments, you need to know several things about your present or planned VoIP configuration. If you don’t want to use VoIP Assessor’s default settings, you might want to set or select the following VoIP Assessor parameters to run simulated voice traffic on an IP network that closely resembles the real VoIP traffic you’ll be running:

- Six different codecs, incorporating compression algorithms, data rates, and datagram sizes.
- Whether packet loss concealment (PLC) is enabled for the G.711 codecs.
- Voice datagram sizes.
- The ability to use silence suppression.
- A jitter buffer and its size.
- Any additional, fixed delay values that apply.
- Quality of Service (QoS).

Let’s look at each of these parameters in detail.

Codecs

In audio processing, a codec (which stands for "compressor/decompressor") is the hardware or software that samples the sound and determines the data rate. There are dozens of available codecs, each with different characteristics. VoIP Assessor lets you perform voice over IP assessments using six call scripts representing the six most common codecs.

Codecs have odd-looking names that correspond to the name of the ITU standard that describes their operation. The codecs named G.711u and G.711a convert from analog to digital and back with high quality. To do this, however, takes a fair amount of bandwidth.

The lower-speed codecs, G.726-32, G.729, and those in the G.723.1 family, consume less network bandwidth. Low speed codecs impair the quality of the audio signal much more than high-speed codecs, however, because they compress the signal with lossy compression. Fewer bits are sent, so the receiving side does its best to approximate what the original signal sounded like.

Packet loss concealment (PLC) is an additional option if you’re using the G.711u or G.711a codecs. PLC techniques reduce or mask the effects of data loss during a VoIP conversation. When PLC is enabled, VoIP Assessor assumes that the quality of the conversation would be improved; this improvement is factored into the MOS estimate calculation if any data is lost. PLC makes the codec itself more expensive to manufacture, but does not otherwise add delay or have other bad side-effects. If the VoIP equipment you plan to purchase uses PLC, be sure to enable it in your assessment.

In the table of codecs on the next page, the “Packetization Delay” column refers to the delay a codec introduces as it converts a signal from analog to digital. Packetization delay is included in the MOS estimate, as is the “jitter buffer delay,” the delay introduced by the effects of buffering to reduce inter-arrival delay variations.

Real bandwidth consumption by VoIP calls is higher that it first appears. The G.729 codec, for example, has a data payload rate of 8 kbps. Its actual bandwidth usage is higher than this, though. When sent at 20ms intervals, its payload size is 20 bytes per datagram. To this add the 40 bytes of RTP header (yes, the header is bigger than the payload) and any additional layer 2 headers. For example, Ethernet adds 18 more bytes. Also, there are two concurrent G.729 RTP flows (one in each direction), so double the bandwidth consumption you’ve calculated so far. The “Combined Bandwidth” column in the table shows a truer picture of actual bandwidth usage for some common codecs.

It’s worth observing in the table below that both G.723.1 codecs result in calls of only “Acceptable” quality at best. Their theoretical maximum MOS is below the 4.03 value needed to be considered “Good.”
## Determining Payload Sizes

Each call script lets you “Override delay between voice datagrams.” This option is recommended only for advanced users, and it’s not likely you’ll need to set it. However, here’s an explanation of what happens when you choose to override the delay associated with the codec you’re using:

The delay between datagrams determines the payload size to be used in the simulated VoIP calls. VoIP applications break voice data into chunks based on the amount of time between successive datagrams. Each call script uses a value appropriate to its codec: the faster codecs use 20 ms and the G.723.1 codecs use 30 ms. This means, for example, that every 20 milliseconds, the VoIP application adds a header to any voice data it has received and sends the datagram. Smaller times mean that the header-to-payload ratio is larger: more, smaller datagrams are sent, which increases processing overhead. Longer times mean that fewer, larger datagrams are sent.

Unless you’re an equipment manufacturer, you’re better off leaving the datagram sizes at their default values. These values were selected because they match those of the codec being emulated and allow for realistic simulation of VoIP datagram traffic.

Some VoIP equipment refers to the delay between voice datagrams as the “speech packet length.” For example, at 64 kbps, a “20 millisecond speech packet” implies that the sending side creates a 160-byte datagram payload every 20 ms. A simple equation relates the codec speed, the delay between voice packets, and the datagram payload size:

\[
\text{Datagram payload size (in bytes)} = \frac{\text{Codec speed (bits/sec) x packet delay (ms)}}{8 \text{ (bits/byte) x 1000 (ms/sec)}}
\]

In this case:

\[
160 \text{ bytes} = \frac{(64000 \times 20)}{8000}
\]

For a given data rate, increasing the delay increases the datagram size, since datagrams are sent less frequently. A packet delay of 30 ms at a data rate of 64000 bps would mean sending 240-byte datagrams.

### Silence Suppression

Call scripts have the option to include silence suppression. Performed by the codec, VoIP silence suppression means that no data is sent on the network during periods of call silence (that is, when no one is talking during the call). If you set were to set the silence suppression option to use a voice activity rate of 50%, it means that data is being sent during 50% of the simulated call’s duration.

When doing your initial assessments, we recommend turning silence suppression off, to provide a more intense assessment of the

<table>
<thead>
<tr>
<th>Codec</th>
<th>Default Data Rate</th>
<th>Default Datagram Size</th>
<th>Packetization Delay</th>
<th>Combined Bandwidth for 2 Flows</th>
<th>Default Jitter Buffer Delay</th>
<th>Theoretical Maximum MOS</th>
</tr>
</thead>
<tbody>
<tr>
<td>G.711u</td>
<td>64.0 kbps</td>
<td>20 ms</td>
<td>1.0 ms</td>
<td>174.40 kbps</td>
<td>2 datagrams (40 ms)</td>
<td>4.40</td>
</tr>
<tr>
<td>G.711a</td>
<td>64.0 kbps</td>
<td>20 ms</td>
<td>1.0 ms</td>
<td>174.40 kbps</td>
<td>2 datagrams (40 ms)</td>
<td>4.40</td>
</tr>
<tr>
<td>G.726-32</td>
<td>32.0 kbps</td>
<td>20 ms</td>
<td>1.0 ms</td>
<td>110.40 kbps</td>
<td>2 datagrams (40 ms)</td>
<td>4.22</td>
</tr>
<tr>
<td>G.729</td>
<td>8.0 kbps</td>
<td>20 ms</td>
<td>25.0 ms</td>
<td>62.40 kbps</td>
<td>2 datagrams (40 ms)</td>
<td>4.07</td>
</tr>
<tr>
<td>G.723.1 MPMLQ</td>
<td>6.3 kbps</td>
<td>30 ms</td>
<td>67.5 ms</td>
<td>43.73 kbps</td>
<td>2 datagrams (60 ms)</td>
<td>3.87</td>
</tr>
<tr>
<td>G.723.1 ACELP</td>
<td>5.3 kbps</td>
<td>30 ms</td>
<td>67.5 ms</td>
<td>41.60 kbps</td>
<td>2 datagrams (60 ms)</td>
<td>3.69</td>
</tr>
</tbody>
</table>
network. This lets you see behavior in the worst case, where there’s a lot of concurrent back-and-forth talking. To do this, set the voice activity rate to 100%.

**Jitter Buffers**

To minimize call disruptions from delay and jitter, VoIP phones and gateways typically use jitter buffers. A jitter buffer can be either frame-based or absolute: a frame-based jitter buffer will hold a given number of voice datagrams, while an absolute jitter buffer is based on time. For example, a frame-based jitter buffer might hold 2 datagrams, buffering them until a segment of the voice transmission can be reassembled to reduce inter-arrival time variability. An absolute jitter buffer, on the other hand, might be set to 43 ms, and given a typical 20 ms speech frame size, could hold 2 speech frames and allow for an extra 3 milliseconds of variability.

It’s the nature of IP networks to exact a trade-off. Buffering adds delay while smoothing out variability; therefore, adding delay can offset the positive effects of using jitter buffers. A goal of VoIP network tuning is to minimize jitter buffer sizes while maintaining call quality.

By default, all VoIP assessments emulate a frame-based jitter buffer of 2 datagrams. You can configure buffers based either on time (in milliseconds) or on number of datagrams. The supported range of values is 10-1600 ms for an absolute jitter buffer, and 1-8 for a buffer configured in number of voice datagrams.

VoIP Assessor’s jitter buffer doesn’t smooth out jitter. Instead, it provides a more accurate MOS estimate by better accounting for datagrams that would have been lost due to underrun or overrun of the jitter buffer. In addition, the delay incurred by using a jitter buffer is factored into the MOS.

We suggest doing your initial VoIP assessments with jitter buffer sizes that match those being used in your planned VoIP equipment, if that information is known.

**Additional Fixed Delay**

This parameter lets you add a delay value to each VoIP call. Any fixed delay that you add here should come from a known, constant source. For example, if you’re emulating equipment that adds 10 milliseconds of delay to each VoIP datagram sent, enter 10 ms. The range of accepted values is 0 to 300 ms.

**Quality of Service**

Using a prioritization or QoS scheme can make a noticeable difference in call quality, particularly if you’re running VoIP on a crowded enterprise network. VoIP Assessor’s optional QoS parameter applies the QoS used in typical VoIP implementations to the simulated VoIP traffic. VoIP Assessor’s QoS option has been optimized for voice traffic, which means that each voice datagram

- is assigned a priority that makes it unlikely to be queued or dropped, or
- is flagged to receive the lowest possible delay.

When you select QoS, VoIP Assessor sets the IP Type of Service (TOS) precedence bits in the IP header, using the “CRITIC/ECP” setting. These same bits are known as the Expedited Flow (EF) setting in the Differentiated Services (DiffServ) standard [6,7]. Unfortunately, the setting of the DiffServ or TOS bits is not supported by every TCP/IP stack. Check the documentation for the endpoint and its operating system to make sure the setting of these bits is honored.

If you haven’t yet tried applying QoS to VoIP traffic, you might want to set up a few call groups that use QoS and a few that don’t. Depending on the routers’ ability to process requests for QoS, you should see a difference in the groups’ MOS values when you compare the results.

By default, none of VoIP Assessor’s call scripts uses a service quality. To use QoS, you must add a new call script with the service quality option enabled.

**Factors Included in Call Quality Estimation**

Once an assessment has run, you can generate two different reports of the results: an Executive Summary Report, and a Complete
Report. Each report gives an overall assessment of call quality for the entire network; the Complete Report adds the individual results for each call group.

To determine the relative quality of each simulated VoIP call made during a VoIP Readiness Assessment, VoIP Assessor measures delay, jitter, and lost data.

**Delay**

End-to-end delay, the time it takes to get data across the network, is the primary indicator of the “walkie-talkie” effect in a VoIP call. People are used to talking at the same time in their telephone conversations, or at least hearing the other person speak in real time. Most listeners notice when the delay preventing their words from reaching the other person is more than about 150 ms; when it exceeds 200 ms, they generally find it disturbing and describe the voice quality as poor.

The end-to-end delay is actually made up of five components:

- **Propagation delay**: the time to travel across the network from end to end. For example, the propagation delay between Singapore and Boston is much longer than the delay between New York and Boston.
- **Transport delay**: the time to get through the network devices along the path. Networks with many firewalls, many routers, congestion, or slow WANs introduce more delay than an overprovisioned LAN on one floor of a building.
- **Packetization delay**: the time for the codec to digitize the analog signal and build frames – and undo it at the other end. The G.729 codec has a higher packetization delay than the G.711 codecs because it takes longer to compress the signal.
- **Jitter buffer delay**: the delay introduced by the receiver as it holds one or more datagrams to reduce variations in arrival times.
- **Fixed delay**: an additional user-configured delay that can be specified when setting up an assessment.

The combined value of propagation delay and transport delay is termed “network delay” in VoIP Assessor. Packetization delay and the jitter buffer delay are constants for any given call group.

Measuring response time (that is, round-trip delay) and dividing the resulting time measurement by two isn’t always a good approximation of one-way network delay. Response time hides assumptions about the symmetry of the paths between two endpoints. The two RTP streams in a VoIP call can take different paths through an IP network.

Endpoints therefore calculate network delay explicitly, rather than just taking the round-trip time and dividing it in half. The endpoints start with flows similar to those used by the Network Time Protocol (NTP). NTP generally has an accuracy of around plus-or-minus 200 ms. Our design for giving good MOS feedback called for clock precision of about plus-or-minus 1 ms, which led us to design more precise algorithms for software-based clocking. We’ve recently applied for patents on these new algorithms.

Endpoints maintain local copies of their clocks because each endpoint can be included in many simultaneous call groups. Also, the internal clocks in every different operating system and computer platform seem to be a little different, and the clocks drift apart over time.

The endpoints maintain virtual (software) clocks for each partner involved in a VoIP test. These virtual clocks consist of the offset between the microsecond clocks maintained by the two endpoints. A high-resolution clock, the microsecond clock is maintained independently of the operating system’s system clock.

The endpoints paired with each other in a call group compare their respective views of the clocks prior to the start of each test and periodically during a test run. They also measure clock synchronization and drift between test runs, to establish a track record for the expected delay.

Our one-way delay algorithms have proven robust, in measurements with thousands of endpoint pairs. We’ve also verified their effectiveness in testing with stratum 1 GPS timeservers.
Jitter

As simulated calls run during an assessment, the endpoints calculate jitter, a factor known to adversely affect call quality. Jitter, also called delay variation, indicates the differences in arrival times among all datagrams sent during a VoIP call.

When a datagram is sent, the sender (one of the endpoints) gives it a timestamp. When it’s received, the receiver adds another timestamp. These two timestamps are used to calculate the datagram’s transit time. If the transit times for datagrams within the same call are different, the call contains jitter. In a video application, jitter manifests itself as a flickering image, while in a telephone call, its effect may be similar to the effect of lost data: some words may be missing or garbled.

The amount of jitter in a call depends on the degree of difference between the datagrams’ transit times. If the transit time for all datagrams is the same (no matter how long it took for the datagrams to arrive), the call contains no jitter. If the transit times differ slightly, the call contains some jitter. Jitter values in excess of 50 ms probably indicate poor call quality. They provide a short-term measurement of network congestion and can show the effects of queuing within the network.

When jitter is detected, jitter totals are factored into the Mean Opinion Score estimate and also reported separately. You can specify what levels of jitter are acceptable on a network by configuring result ranges at the VoIP Assessor console.

Lost Data

In VoIP Readiness Assessment reports, VoIP Assessor includes statistics on lost datagrams, expressed as a percentage of all data sent in the relevant calls. For example, in charts indicating lost data by call group, lost data is expressed as a percentage of all data sent between the endpoints in the call group over the course of the entire assessment. Other charts might show data loss as a percentage of data sent at a certain time of day, averaged over the course of all days in the assessment.

VoIP packets are sent using RTP, the real-time transport protocol. Although every RTP datagram contains a sequence number, there isn’t enough time to retransmit lost datagrams. Any lost datagram impairs the quality of the audio signal.

There are two primary reasons why RTP datagrams are lost in a data network:

- there’s too much traffic, so datagrams are discarded during congestion, or
- there’s too much delay variation, so datagrams are discarded because they arrive at the listener’s jitter buffer too late or too early.

There are a couple of patterns to datagram loss. The simplest “pattern” is when there’s a more-or-less random loss. There’s general, consistent congestion in the network, so one or two datagrams are lost occasionally. The other loss pattern is when packets are lost in bursts, say five or more at a time. Humans perceive that bursts of loss impair audio quality much more than general, random loss.

When a datagram is lost during a VoIP transmission, you can lose an entire syllable or word in a conversation. Obviously, data loss can severely impair call quality. VoIP Assessor therefore includes data loss as a call quality impairment factor in calculating the MOS of each simulated VoIP call.

To measure data loss, one computer in each call group keeps track of how many bytes of data it sent. The sending endpoint reports to the receiving endpoint how many bytes it sent, and the receiver compares that value to the amount received to determine lost data.

As mentioned in our discussion of codecs, above, having PLC enabled for the G.711 codecs improves call quality in the presence of lost data (compared to what the MOS would be if PLC were disabled). Always enable PLC in your assessments, if it’s supported by the VoIP equipment you plan to deploy.

Getting Started

The first thing you need to do to get started running a VoIP Readiness Assessment is to install Performance Endpoints on the appropriate computers in the data network. Then take the following steps:
• **Design** the assessment schematically in the Design view at the VoIP Assessor console. Choose the computers where the simulated VoIP traffic will run and connect them using VoIP connectors. See “Step 1: Design the Assessment” on page 12 for more information.

• **Schedule** the assessment in the Schedule view. See “Step 2: Schedule the Assessment” on page 14 for more information.

• **Verify** that the computers you’ve selected can be reached, and that they have Performance Endpoint software installed. See “Step 3: Verify the Assessment” on page 15 for more information.

• **Run** the assessment. See “Step 4: Run the Assessment” on page 15 for more information.

• **Report** the results. Generate a report summarizing VoIP quality and network readiness.

But first, take some time to plan the assessment.

**Planning an Assessment**

Before you design your first VoIP Readiness Assessment, you need to do some research on the data network and on the planned VoIP implementation.

**Understanding Call Volume**

First, look at existing network documentation to find peak and average usage statistics. How many calls are likely at the same time? What is their duration?

Telephone records and the current PBX call volume reports are a good source of data about the likely call volume a network will have to handle. What percentage of the calls occur within each site? What percentage occur within the organization, from site to site? How many calls go to and from the outside world?

For example, we have eight major sites in NetIQ. From our development site in Raleigh, we rarely call our sales offices in Japan and Europe. It makes sense to set up just one call from Raleigh to Japan and from Raleigh to Europe. An assessment generates several calls an hour, although we probably make less than one call a day between these sites. Part of an assessment is to make sure that you can get a connection and make a toll-quality call any time you want one, so testing throughout the day is fine. We call less rarely from Raleigh to Portland, so we would define maybe two calls between those sites. Finally, we make many calls to Houston and San Jose, so we would define ten simultaneous calls.

A VoIP assessment is not a stress test; remember, you’re running on a production network. Test with an approximation of the average call volume during work hours, as opposed to the peak call volume. There’s a nice “failure mode” though, that is easily observed. If the data network is already heavily loaded with existing application traffic, and then you add VoIP traffic, it’s the VoIP traffic that “breaks first” – it will show high delay, jitter, packet loss, or some combination. This is readily seen during the verification step of an assessment. If the VoIP verification shows “green,” the additional VoIP traffic will probably have no adverse effect on the other application traffic. However, if the VoIP verification shows “red,” you may or may not be affecting the other traffic – but you know already that the network resources are stretched too thin.

**Understanding the VoIP Equipment**

Next, check the vendor data sheets for the VoIP equipment you’re considering, to answer the following questions:

- What type of codec will you be using?
- Will this codec use silence suppression or packet loss concealment?
- Can the network currently support QoS for VoIP?
- What size jitter buffer will be used?
- Are there any additional sources of delay to be considered?
- How many simultaneous calls can be supported?

You’ll want to emulate these factors as closely as possible when you configure assessments.

Most IP phones support more than one codec. The codec to be used is negotiable, and is usually centrally managed. When an IP phone is first plugged in, it downloads a list of which
codec to try first, which to try second, and so on.

Does bandwidth consumption force you to use a codec other than G.711? It may be that the lower call quality is good enough in your setup. Also, silence suppression and PLC make telephones more expensive and is rarely supported today. If your phones don’t support these, leave these choices unclicked when setting up the VoIP assessment.

Where to Install Endpoint Software

Before you set up an assessment, you must install Performance Endpoint software on every computer you plan to include in assessments.

Put the endpoint software in the right locations, on the right computers, to get a representative assessment. But, what are the right locations and what are the right computers? Let’s start with the location.

Look at the major offices in your organization, among which the majority of the calls are made today. For example, in NetIQ, we have eight major offices: four development sites and four key sales offices. Put at least two endpoints at each site. Design the assessment to go from each site to each of the others, and define a call group between the two endpoints in each office or site. In our case, with eight offices, we’d define $8+7+6+5+4+3+2+1$ call groups in the VoIP Assessor, a total of 36 call groups.

When connecting within a site, look at the network topology of the site to see what will give the most telltale assessment. If you have multiple, interconnected switches, be sure to put the endpoints on different switches. If your sites have thousands of users and multiple buildings, consider an endpoint for each building or for each division. Place endpoints in the spots where you suspect there will be problems. Do you have a mishmash of ISDN and T1 WAN connections that are stretched to their capacity? These links are better candidates for including in the assessment than a modern site where everyone has a 100 Mb LAN connection to their desktop.

Within a site, your endpoint placement should hinge on the VoIP topology in your plans.

Where are the two ends of the IP connection? Do you plan to continue with analog telephones on the desktops, then VoIP over the WAN? If so, install the endpoint software in a computer adjacent to the PBX or VoIP gateway. Alternatively, do you plan VoIP from desks within a site or campus, connected to the external public switched telephone network (PSTN)? In this case, put endpoints on a few desks, connected to endpoints adjacent to the gateway.

Install a few endpoints on either side of a WAN link, which should be part of any assessment. Install endpoints on either side of a firewall and set up call groups to test how well the firewall will handle VoIP traffic.

Finding the right computer for endpoint software is easier. The endpoint need not be installed on a dedicated computer. Do you have a computer available with some extra capacity, say a domain name server, a DHCP server, a print server, or a remote terminal server? These are good candidates for endpoints, as long as they are used for less than 30 call groups. Otherwise, put the endpoint software on an extra computer for the duration of the assessment. Almost any Pentium II or above is sufficient for a Windows or Linux endpoint; most SPARC or Ultra CPUs from the past few years will run a Solaris endpoint well.

Step 1: Design the Assessment

In VoIP assessments, endpoints are computers with NetIQ Performance Endpoint software installed. The Design view lets you choose which endpoint computers to include in an assessment, and which simulated calls they’ll receive.

To design an assessment, first create some endpoints. The Create an Endpoint dialog lets you enter an IP network address to represent a VoIP-enabled computer on the network. Newly-created endpoints are added to the Endpoint List to the right of the diagram.
Next, you connect the endpoints using VoIP connectors. Click the VoIP Connector tool, click on an endpoint, drag the connector to a second endpoint, and click again. The Create a VoIP Connector dialog lets you select key parameters for the voice traffic that will be sent between the endpoints, including the type of codec emulated and the number of simultaneous calls represented. The combination of two endpoints and a VoIP connector makes a call group, the basic unit of a VoIP assessment.

Creating an Endpoint

When you fill in the fields provided on the Create an Endpoint dialog, an icon for the new endpoint appears on the network diagram. Here are the key fields.

Endpoint Name
Give the endpoint a name that makes it easy to identify in the reports.

Network Address
This is an endpoint computer’s network address, or a set of addresses assigned to that computer’s network interface card (NIC). Enter a DNS hostname, such as www.netiq.com, or the IP address of the endpoint computer in dotted notation, such as 135.25.25.5.

Address Type
The network addressing scheme that applies to this endpoint computer’s NIC.

A single computer can be assigned multiple IP addresses for advanced assessments. Choose Single IP Address or Range of Virtual IP Addresses from the list. Enter the address(es) in the Network Address or From and To fields.

The option to enter a range of virtual IP addresses is included in case you want to create tests with many calls using different addresses, but you only want to use a few computers to represent those calls.

Setup Address (optional)
The network address at which the VoIP Assessor console contacts an endpoint with setup information. The console sends endpoint computers information about how to simulate VoIP calls for the assessment, and it can send these setup flows over a different connection from the one specified above in the “Network Address” field. This is the address by which VoIP Assessor identifies the endpoint containing the virtual addresses, if any are used.

Contact Endpoint Using Network Address
Instructs the VoIP Assessor console to use the endpoint’s IP address to contact it when it acts as Endpoint 1 (“From Endpoint”) in a call group. If this box is checked, you don’t need to enter a Setup address.
Creating a VoIP Connector

By connecting two endpoints, VoIP connectors create call groups that emulate VoIP calls between endpoint computers. The Create a VoIP Connector dialog lets you select the parameters that match your VoIP application.

From Endpoint
The endpoint that will act as the initiator of the call in this call group. In the Call Groups table, designated as Endpoint 1.

To Endpoint
The endpoint that will act as the recipient of the call in this call group. In the Call Groups table, designated as Endpoint 2.

Call script
The codec used to send VoIP datagrams, plus any additional parameters that affect the calls that are sent on the network.

Number of concurrent calls
The number of simultaneous calls VoIP Assessor will emulate during assessments including these endpoints. Values must be between 1 and 1,250, inclusive. The default value is 1.

After you create a call group, the Call Count in the lower right of the assessment window shows you how many VoIP calls will be emulated.

Design Tips
Create endpoints to represent the different areas of a network. Include at least one call group that spans a WAN link, assuming your deployment will include one. When creating VoIP connectors, create only one connector for each call script (codec) you want to include between any two endpoints in the assessment. To add more calls, change the “Number of concurrent calls” in the Create a VoIP Connector dialog.

If there’s a firewall between the endpoints in a call group, you need to do some extra configuration, both in VoIP Assessor and at the firewall itself.

Firewall Settings
If there’s a firewall on the network between the endpoint computers, you’ve probably already configured it to accept RTP flows for VoIP traffic. You also need to open a port to allow endpoints in each call group to receive setup information and return results. The console communicates with the endpoints using bi-directional TCP over port 10115.

Enter a Reporting port for the endpoints to use when reporting results back to the VoIP Assessor console, or leave the default at TCP port 10116. Choose AUTO to let the endpoints choose the port dynamically. Make sure the port you choose is open to bi-directional TCP traffic at the firewall.

Enter a Call Traffic port or range of ports to use if the firewall has been configured to pass VoIP calls through one particular port. The endpoints need to be able to send simulated VoIP traffic to each other through the firewall. Within any range of ports you set, the endpoints will only use even-numbered ports (which is what real VoIP RTP streams use). The default, AUTO, lets the endpoints choose the port dynamically.

The port you use as a Call Traffic port must be opened at the firewall to bi-directional RTP traffic. All simulated VoIP traffic that VoIP Assessor sends over the network is bi-directional, to emulate the bi-directional nature of voice over IP.

If the firewall does not allow RTP traffic, open the port to allow UDP traffic.

After you set or change the reporting port, you must stop and restart the VoIP Assessor Scheduler service.

Step 2: Schedule the Assessment

To determine when the assessment will run, create a schedule. Scheduling options include the following:

- The length of time the assessment will run
- The duration of each set of calls
- The interval between each set of calls.

The minimum interval between calls is 15 minutes. Use the minimum interval to emulate heavier VoIP traffic. Or increase the interval to measure the call quality when VoIP traffic is lighter.
The endpoints need some time to report the results of the simulated calls to the console. Therefore, the minimum length of time between the duration and interval you set must be five minutes. For example, if you set a call duration of 20 minutes and a calling interval of 22 minutes, you’ll be prompted to increase the interval or decrease the duration of the calls. In this case, you could simply increase the interval to 25 minutes.

When you have made your selections, validate the schedule to apply your settings to the current assessment and commit them to the database.

**Scheduling Tips**

A good length for an assessment is about seven days, enough to get plenty of data, from different days of the week when network traffic conditions are different. However, for your first assessment, we’d suggest that you run for a few hours and then scan your results. If it appears that the assessment is providing meaningful results and including enough representative endpoints to cover the network, reset the schedule for a longer time period. Likewise, start with fewer calls, and add more after a successful test assessment. You can easily edit VoIP connectors in the Design view to add more calls.

The schedule interval and the number of concurrent calls determine how heavy the simulated VoIP traffic will be. To obtain more data points, choose a shorter interval. This helps increase the chance that you identify a drop in quality due to regular spikes in data traffic. For instance, if the data traffic always spikes on the hour, but you never have VoIP test calls running at that moment, you may lose the fact that VoIP quality may drop then. You should have some idea of what peak and average utilization on the network will be; if not, do some traffic analysis to find out.

**Step 3: Verify the Assessment**

Verification is required before a VoIP assessment can begin. The console contacts each of the selected endpoint computers and sends them a brief schedule and one call script.

Verification is like a mini-assessment. You get some preliminary results after you verify your assessment, and you see right away whether you need to make any changes to the assessment design—any errors that occurred are indicated.

If errors occur during verification, you’ll know it by the warning icon that appears in the Call Groups table. Look in the error log to see what happened and which endpoints were affected.

- If Performance Endpoint software wasn’t detected by one of the computers in the assessment, install the software.
- Return to the Design view and make the necessary changes to the assessment configuration to avoid repeating the error.

**Step 4: Run the Assessment**

In the Run view, you can start the assessment and determine how it’s proceeding. The status is continually updated. If an error occurs, the error log shows what happened. You then have the option to stop the assessment and change configuration parameters before starting it again.

Although the Run view lets you stop an assessment, VoIP Assessor’s scheduler component will also stop the assessment when the time period you entered in the Schedule view has passed.

The lower half of the Run view contains a table of preliminary results for each call group.

While the assessment is running, your presence is not required. However, you should plan to check the assessment periodically to make sure no errors have occurred that might interfere with results. The Run view lets you keep tabs on the assessment and indicates errors as they occur.
Step 5: Report Results

As soon as an assessment runs long enough to show some results, you can generate a report of these results.

VoIP Assessor offers two pre-configured report types:

- **Executive Summary Report** — provides a summary, with relevant charts and graphs, of the results that most directly affected the overall quality of VoIP calls on the network. This report is usually about six pages long.

- **Complete Report** — provides information, charts, and graphs about every type of data collected during the assessment. Complete Reports take varying amounts of time to generate, depending on the number of calls and call groups in the assessment.

For an assessment of 6 call groups emulating 24 calls every 15 minutes for 7 days, the report takes about 10 minutes to generate on a 600 MHz computer with 256 MB of RAM. However, an assessment running 1250 calls every 15 minutes for 7 days will take over 9 hours to generate! For larger assessments, plan to let VoIP Assessor generate the report while you’re away from the console.

When you generate a report, VoIP Assessor invokes Microsoft’s Word and Excel programs to interact with its database and produce charts, graphs, and detailed explanations of the data in a formatted report.

Although you can create custom reports by gleaning data from the database using Microsoft Access, another option is to edit the Complete Report in Microsoft Word, removing the sections you don’t need.

Look at the sample report that ships with VoIP Assessor to get acquainted with the kind of data you can get from your VoIP Readiness Assessment.

Reports are more useful if you assign helpful endpoint names to the endpoints. You’ll notice that by default, the reports label the results according to the call group names, which are combinations of the endpoints’ names and call scripts. Another option is to assign a Connector Comment to each VoIP connector to achieve the same purpose: identifying the call groups in reports. Names that describe the locations of the endpoints are perhaps the most helpful.

Analyzing Results

VoIP Assessor reports aim to be self-explanatory. However, it’s helpful to understand how the various charts and tables break out and present the results.

**Call Quality Charts**

Just like delay, jitter, and lost data, call quality is measured in units. In the case of call quality, the units are points on the 5-point MOS scale. Therefore, in bar charts showing call quality, the bars show the average MOS of all calls made between the endpoints in a call group. In addition, the bars are further broken down into call-quality mappings to show the percentage of all calls that were rated as having Good, Acceptable, and Poor quality or were Unavailable.

On charts that break out specific impairment factors by day or by hour, a line graph of that factor’s values is superimposed over a bar graph. For example, the line graph of **Delay by Day** gives you a quick overview of the lost data results averaged for each day of the assessment (see below). The bars show what percentage of the lost data values fell into the Good, Acceptable, and Poor ranges.
Delay by Day

To take a closer look at the actual data, double-click any chart to launch Microsoft Excel. You then have access to the spreadsheet containing your data.

Data Tables

In the reports, bar charts are paired with data tables so that you can quickly scan the results and see the specific values. Different colors in the bar charts provide an overview of the data, while the tables provide the specifics.

Most values shown are based on averages. For example, when you look at the charts that break out call quality by time of day (“Call Quality by Hour,” for example), you’re seeing MOS averages for all simulated calls that were made during a certain hour of the day over the course of the entire assessment.

Summary

With the right tools, the process of deploying VoIP successfully on a data network can become a straightforward decision tree.

Run a VoIP Readiness Assessment. Look carefully at call quality, from a single call to the maximum number of expected calls at peak network usage. Compare calls across a range of locations.

If the call quality is good and the other traffic is relatively unaffected, great – it avoids lots of complexity. Start the VoIP deployment.

If the call quality is not acceptable, determine what the problems are and where they’re located. What factor influenced the poor quality the most: one-way delay, jitter, lost data, or a combination of all three? Can a simple change in the VoIP configuration options, such as the choice of codec, improve the call quality?
quality sufficiently? Where are the most likely bottlenecks?

Now, look at the costs of making the required network improvements. Choices include adding more bandwidth, upgrading or replacing existing equipment, changing the network design for greater efficiency, reconfiguring or tuning the network for QoS, or a combination of these.

These choices are only the start of a decision tree for a network administrator, because the costs of these different choices are not equal. Adding more bandwidth may be a recurring expense; upgrading the hardware may be a capital expense; and QoS may appear to be free, but it usually has a high cost in personnel time.

Analyze the costs in as much depth as you can and decide whether you want to proceed with network changes. It’s an iterative process of making the most cost-effective improvements a step at a time, then repeating the VoIP Readiness Assessment to see if you’re reaching your call-quality goals.

If your cost estimates for preparing the data network for VoIP appear too high, this is a good time to take another look at your VoIP deployment plan. By this point, you’ll understand better what the deployment will require, so you’ll have some choices:

- you can decide how to budget costs intelligently at the right time in the future,
- you can increase your current budget and proceed – considering this a suitable long-term investment, or
- you can approach VoIP as a staged deployment, taking some steps now and saving some steps for later.

About the Authors

Susan M. Pearsall, Ph.D., is a senior writer with NetIQ Corporation. She can be reached at susan.pearsall@netiq.com. John Q. Walker, Ph.D., is the director of network development and a co-founder of Ganymede Software, which joined NetIQ Corporation in 2000. He can be reached at johnq@netiq.com.

Acknowledgments

This paper was suggested and designed by Kim Shorb and Dave Zucker. Gracious thanks to the readers who helped improve this paper: Jeff Aldridge, Lynne Attix, Bruce Dean, Jeff Dozer, and Steve Joyce.

For More Information

Copyright Information

NetIQ Corporation provides this document “as is” without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of merchantability or fitness for a particular purpose. Some states do not allow disclaimers of express or implied warranties in certain transactions; therefore, this statement may not apply to you. This document and the software described in this document are furnished under a license agreement or a non-disclosure agreement and may be used only in accordance with the terms of the agreement. This document may not be lent, sold, or given away without the written permission of NetIQ Corporation. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, or otherwise, with the prior written consent of NetIQ Corporation. Companies, names, and data used in this document are fictitious unless otherwise noted. This document could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein. These changes may be incorporated in new editions of the document. NetIQ Corporation may make improvements in and/or changes to the products described in this document at any time.

© 1995-2002 NetIQ Corporation, all rights reserved.

U.S. Government Restricted Rights: Use, duplication, or disclosure by the Government is subject to the restrictions as set forth in subparagraph (c)(1)(ii) of the Rights in Technical Data and Computer Software clause of the DFARs 252.227-7013 and FAR 52.227-29(c) and any successor rules or regulations. AppManager, the AppManager logo, AppAnalyzer, Knowledge Scripts, Work Smarter, NetIQ Partner Network, the NetIQ Partner Network logo, Chariot, End2End, Pegasus, Qcheck, OnePoint, the OnePoint logo, OnePoint Directory Administrator, OnePoint Resource Administrator, OnePoint Exchange Administrator, OnePoint Domain Migration Administrator, OnePoint Operations Manager, OnePoint File Administrator, OnePoint Event Manager, Enterprise Administrator, Knowledge Pack, ActiveKnowledge, ActiveAgent, ActiveEngine, Mission Critical Software, the Mission Critical Software logo, Ganymede, Ganymede Software, the Ganymede logo, NetIQ, and the NetIQ logo are trademarks or registered trademarks of NetIQ Corporation or its subsidiaries in the United States and other jurisdictions. All other company and product names mentioned are used only for identification purposes and may be trademarks or registered trademarks of their respective companies.