

Realtime
publishers

"Leading the Conversation"

The Definitive Guide[™] To

Converged Network Management



Ken Camp

Chapter 5: Key Steps in VoIP Deployment and Management for the Enterprise	93
Network Readiness Assessment	93
Ensuring Network Readiness for Converged Services	93
Plan to Succeed—Don't Fail to Plan	94
Analysis.....	95
Planning	95
Testing.....	95
Acquire the Right Resources	95
Look to the Future—Think Long Term	96
Understanding Existing Voice Needs	96
One Voice Factor—Compression Methods	97
Voice Call Quality—The PSTN vs. VoIP	99
Network Design Considerations	102
The Performance Envelope.....	103
Choosing Performance Envelope Characteristics to Measure	109
Throughput.....	110
Bandwidth.....	110
Response Time.....	110
CPU Utilization.....	111
Network Segment Utilization	111
Integrity/Reliability.....	111
Packet Loss	111
Jitter.....	111
Delay/Latency	111
Cost	112
Availability	112
Uptime—99.999%	112
Security	112
Manageability	113
Scalability	113
Summary	113

Copyright Statement

© 2007 Realtimepublishers.com, Inc. All rights reserved. This site contains materials that have been created, developed, or commissioned by, and published with the permission of, Realtimepublishers.com, Inc. (the "Materials") and this site and any such Materials are protected by international copyright and trademark laws.

THE MATERIALS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. The Materials are subject to change without notice and do not represent a commitment on the part of Realtimepublishers.com, Inc or its web site sponsors. In no event shall Realtimepublishers.com, Inc. or its web site sponsors be held liable for technical or editorial errors or omissions contained in the Materials, including without limitation, for any direct, indirect, incidental, special, exemplary or consequential damages whatsoever resulting from the use of any information contained in the Materials.

The Materials (including but not limited to the text, images, audio, and/or video) may not be copied, reproduced, republished, uploaded, posted, transmitted, or distributed in any way, in whole or in part, except that one copy may be downloaded for your personal, non-commercial use on a single computer. In connection with such use, you may not modify or obscure any copyright or other proprietary notice.

The Materials may contain trademarks, services marks and logos that are the property of third parties. You are not permitted to use these trademarks, services marks or logos without prior written consent of such third parties.

Realtimepublishers.com and the Realtimepublishers logo are registered in the US Patent & Trademark Office. All other product or service names are the property of their respective owners.

If you have any questions about these terms, or if you would like information about licensing materials from Realtimepublishers.com, please contact us via e-mail at info@realtimepublishers.com.

[**Editor's Note:** This eBook was downloaded from Realtime Nexus—The Digital Library. All leading technology guides from Realtimepublishers can be found at <http://nexus.realtimepublishers.com>.]

Chapter 5: Key Steps in VoIP Deployment and Management for the Enterprise

Different applications have different requirements. The introduction of voice and streaming video onto the existing IP network presents a completely new set of requirements to the operational performance envelope of the network. This chapter will examine the importance of assessing the readiness of the network and fully evaluating design considerations related to delivering integrated service in the enterprise. To address new service integration, the chapter will explore a methodology, called the “performance envelope,” for mapping the characteristics of the corporate network.

Network Readiness Assessment

As businesses delve into VoIP and video solutions in the enterprise, their focus is on business drivers. For many companies, cost reduction is a key business driver for convergence. Cost cannot be the only driver, and for practical purposes, shouldn't be the primary driver. It's important to set reasonable expectations and to fully understand all the business drivers behind service integration. A new application service or an integrated VoIP and CRM solution will require consideration of very different factors than cost reduction alone. It's vital to fully understand the business motivation for converging video, voice, and data so that the project planning and implementation teams can address the true success factors when mapping out project milestones. Communicating the objectives clearly with everyone involved helps maintain a clear view of the expected results.

Ensuring Network Readiness for Converged Services

Ensuring network readiness is a significant task and not to be taken lightly. Over the past 4 or 5 years, both VoIP vendors and systems integrators have learned the hard lesson of failure because of inadequate preparation and planning. The existing data network has been optimized over time to support the existing business requirements. As new applications and services are added, network tuning takes place and a variety of parameters are often tweaked. Historically, these optimization efforts have been driven by packet data applications such as CRM, ERP, and Web services solutions—“normal” IP applications. Voice and video present new challenges as they add streaming, real-time traffic to an existing service network. VoIP call quality requires low latency in the network. Jitter needs to be low, and to avoid complicated jitter buffering requirements, it needs to be consistent. Bandwidth and packet loss need to support the new, integrated services.

Testing and documenting the parameters and characteristics of the existing networks before extensive planning will help ensure the network is capable of supporting VoIP, video, or both. It's important to understand aspects of the traffic needed to support all the different service types that will coexist on the new network. Traffic types, frame size, prioritization schemes for quality of service (QoS), jitter, latency, and packet loss are all crucial factors. Because consistent network performance is so important, it's also prudent to evaluate utilization in the network both at peak and normal times.

If the network will include PSTN gateways for global connectivity, there may be requirements specific to the gateway demarcation between networks. Both the IP network and PSTN need to be considered. The PSTN side of the gateway will have specific trunking requirements. These have traditionally been calculated using Erlang-B formulas. The IP network must support the bandwidth and other network characteristics to ensure a clean handoff of calls from one network to the other.

Erlang-B Calculation

Telephony carrier central offices and enterprise PBX systems are engineered for trunking requirements using the Erlang-B traffic tables and hundred calling seconds per hour, or CCS. 36 CCS would represent a circuit at maximum occupancy with zero idle time. I won't explore the formulas in any depth as we investigate managing the converged services network, but it's worthwhile to understand the complexity and sophistication of traffic engineering in the traditional telephony environment. The formula for determining voice traffic load that's most widely used is:

$$\text{OfferedLoad} = \frac{\text{Calls/Hour} \times \text{AvgMinutes/Call}}{60} \times \text{erlangs}$$

Erlangs are really a simple concept identifying the number of calling minutes per hour a system can handle. The key differentiator is between the traditional circuit world, where minutes equate to busy circuits, against the VoIP world, where minutes of use may not be applicable. In VoIP services, it's really the carrying capacity of the network in packets that's the key indicator.

The enterprise IP network must complement the trunking requirements when connecting to the PSTN. Overlooking details like this may result in performance bottlenecks that impact call quality.

Plan to Succeed—Don't Fail to Plan

Whenever new technologies are implemented in the corporate network, there are some basic planning steps that should not be overlooked. These basics will improve the likelihood of success in the deployment of converged voice and video services.

Although this brief section seems to be focused on the planning and implementation of integrated services, that is the crucial starting point of holistic management of the total integrated services network. When an organization decides to implement VoIP, that decision commonly leads to acceptance of a new, holistic approach to managing the entire service network.

Analysis

Before undertaking a convergence project, fully identify the business drivers for change. It's important to identify the business objectives and set everyone's expectations accordingly. List the benefits you expect to derive from integrating services so that each benefit can be addressed individually. Broad, sweeping objectives such as "save money" aren't measurable. The success of VoIP implementations really depend on advance data gathering and analysis.

The more you can document prior to any implementation work, the greater your likelihood of success. It's important to balance objectives and expectations with analysis of your network's capabilities. The determination that your network can support the new, converged services, or learning that it can't and will require upgrades in routing and switching, will influence both the budget and project timelines.

Planning

There's an old adage that says "people don't plan to fail, they fail to plan." It's vital that your project team be methodical and document every facet of the entire planning process. One widespread complaint heard from both systems integrators and VoIP vendors has been that the assessment and planning has been the weak link in a high percentage of failed projects.

Resources invested in planning, whether spent on a project manager or the hours spent by a technical team evaluating requirements, will bring benefits at cutover time. There is no substitute for preparation and planning. For an enterprise mission-critical service such as voice, failure can be especially painful, both in stress and in business impact. The more meticulous the planning process, the higher the chance of success.

Testing

Once the network requirements have been documented and are fully understood, test to make sure the existing network can support your VoIP requirements. Comprehensive analysis up front helps ensure that you know what to test and what parameters or characteristics your network requires. It's important to approach testing with the specific aim of supporting the fully converged network. Testing VoIP services alone is not enough. Existing applications should be tested while simulating VoIP. There is a hidden danger of integrating VoIP services and doing all the right things to support call quality, only to starve network resources for other mission-critical applications. It is absolutely vital to take a holistic view of all network services and applications when testing.

Acquire the Right Resources

The "right" resources will vary from organization to organization. For some, acquiring resources might mean dedicating a project manager. Most organizations will need some training to bring existing staff up to speed on VoIP resources. For many companies, this will entail bringing in a VAR partner or systems integrator to bolster the expertise available. There is no shame in admitting you need expertise from outside and working with a consulting partner. Leverage expertise everywhere you can.

Look to the Future—Think Long Term

For several years, VoIP has been viewed somewhat as the final objective. It's important now to recognize that this isn't the case. VoIP is a foundation or building block technology. It's important to look at scalability in terms of users and raw numbers, but that's only scaling vertically. An earlier chapter looked briefly at the Software Oriented Architecture (SOA) and Software as a Service (SaaS) concepts. This guide has also talked briefly about emerging characteristics of unified communications such as presence and availability. These are horizontal scalability factors.

To future-proof your network and ensure it can continue to grow to support the enterprise business, a holistic approach that scales both vertically and horizontally will help enable future growth at an acceptable pace. It will enhance readiness at anticipating that next new application requirement that lies just beyond the horizon now. The one you don't yet know about. The VoIP integration project presents an opportunity to review the overall future capacity of your network.

Understanding Existing Voice Needs

Once you've developed all your business needs and determined that VoIP service integration is the right business move, you need to gather data to fully understand voice services requirements. In telecommunications, there are patterns of calling that can be identified. Call usage reports from either the existing PBX or the incumbent telephony services provider offer a wealth of information. Some variation of these reports should be readily available from the current telephony technology, whether it's carrier-based Centrex-like service or a locally administered PBX.

There are several key elements of voice service that need to be well understood. Business telephone systems often focus on the "busy hour," or that hour of the day the most telephone traffic occurs. This is a starting point, but there are many nuances to consider. The busy hour for inbound calls may be different than the busy hour for outbound calls. The aggregated inbound/outbound busy hour may be different than either taken individually. These busy hours may vary on different days of the week.

Most companies also have an identifiable busy day in the average week. And business managers understand the cyclical nature of their respective businesses. The busy month of the year is generally well known in any business sector.

For large enterprises with multiple divisions, or business units, calling patterns may vary and some may have different requirements and calling patterns than other divisions of the corporate headquarters. Flexibility is key to understanding the needs of all business units.

To identify specific business needs, you must understand the calling flows and volumes. Some enterprise calling patterns are very inbound in nature. This might be easily identified by a company using a call center, but other companies establish large outbound call centers as well. The patterns for inbound and outbound calls are often very different. It's vital to know what is required to support your particular business. It's also important to identify upper call volumes, or absolute peak traffic.

Most enterprises are somehow cyclical in nature. When analyzing calling patterns, it's worthwhile to account for several months or even for the past year. Peak periods vary widely by industry. For example, a retail services company may experience much higher call volumes during the Christmas season than any other time of year, whereas a financial services company might expect the highest call volumes at tax preparation time approaches. It's important to take these peak activity periods into account and not simply design the VoIP services to support day-to-day business needs.

As you develop a strategy for integrating VoIP into the network, it's also a good time to recognize that the PSTN has become a mission-critical component of business operations. The PSTN network has been engineered to deliver 99.999 percent uptime. That is what the traditional carriers tout as their benchmark. This service level equates to roughly 5 minutes of downtime per year. Corporate data networks rarely provide this level of reliability. For many organizations, increasing network uptime will provide incentive for network enhancement and redesign requirements during preparation and implementation of a converged services network.

One Voice Factor—Compression Methods

One factor to consider in VoIP deployment is the encoding method used to convert analog human speech into a digital signal for packetized transmission. There are several encoding techniques used in VoIP solutions. Each uses different compression levels. These compression levels affect the bandwidth needed to transmit the voice packets, the quality of the voice, and the processing time. Pulse Code Modulation (PCM), or G.711 encoding, is widely used. It's the standard encoding used in the PSTN, and is generally assumed to provide the best voice quality, although new wideband codecs are now starting to surpass PCM in some implementations. PCM generates a 64Kbps voice stream. For some companies, bandwidth requirements, weighed against quality considerations, might make a different encoding scheme more appropriate. Increasing the compression of the voice signal reduces the overall bandwidth requirement, but voice quality may suffer as a result.

Even though the PSTN is more than 100 years old, voice quality is still often measured by having a group of people listen to sounds in headphones and subjectively rate the quality. These evaluators rate the sound quality on a scale of 1 to 5. This rating, from 1 to 5, is referred to as the *mean opinion score (MOS)*. A rating of 1 roughly equates to the scratchy sound quality of the intercom speaker in a fast food drive through or a warehouse environment. A 5 is the highest rating and is considered “perfect,” or the theoretically highest-grade voice quality achievable. It has always been the target for conducting corporate business calls, but as you'll see in the chart, not actually achieved. A score of 4.4, garnered by the G.711 codec used in the public telephone network, was described in the past as “toll quality” voice. This has been the traditional carriers' target for business-class voice services.

Although the statistical validity of this human sampling may be questionable to some, this approach has been used for many years and is generally accepted globally. Today, MOS is often determined using ITU Recommendation P.800, which details methods for subjective determination of transmission quality. In the real world, the human ear can clearly distinguish between a 4 MOS and a 4.5 MOS.

Table 5.1 provides a comparison of several widely adopted encoding schemes. This table describes the codec types and algorithms used, the sample size, bit rate, and encoding delay for the algorithm used. The last column identifies the MOS for these codecs. When selecting a codec for VoIP implementations, it's important to remember that delay is cumulative from end to end. Encoding delays can become real call quality concerns in a large enterprise network. If the total delay exceeds 250 milliseconds (ms), call quality will suffer. The ITU-T standards (<http://www.itu.int/>) provide extensive details about the operation of these codecs. Vendors and consulting partners will also be a valuable resource in identifying the best approach for each specific implementation.

Encoding Schemes	ITU Codec Descriptor	Sample Size	Bit Rate	Encoding Delay	MOS
PCM	G.711	8 bits	64kbps	<1 ms	4.4
Adaptive Differential Pulse Code Modulation (ADPCM)	G.726	4 bits	32kbps	1ms	4.2
Sub-Band Adaptive Differential Pulse Mode Modulation (SB-ADPCM)	G.722	8 bits	64kbps	4ms	N/A
Low-Delay Code Excited Linear Predictive (LC-CELP)	G.728	40 bits	16kbps	2ms	4.2
Conjugated Structure Algebraic Code-Excited Linear Predictive (CS-ACELP)	G.729/G.729a	80 bits	6kbps/8/kbps	15ms/10ms	4.2
Algebraic Code-Excited Linear Predictive (ACELP)	G.723.1	160 bits	5.3kbps	37.5ms	3.5

Table 5.1: Encoding schemes.

PCM, or G.711, is the method that has been used in the PSTN. It's perhaps the most widely used codec in VoIP systems because it provides a well-known call quality rating. Every VoIP equipment manufacturer supports G.711. G.726, known as Adaptive Differential Pulse Code Modulation (ADPCM), may reduce the bandwidth requirements by 50 percent while only posing minimal degradation (.2 of a point) on the MOS rating.

G.722 is mostly used in FM radio, but is included here because FM radio quality is a good audio fidelity comparison. It's rarely used in VoIP solutions. G.722.1 offers lower bit-rate compressions. A more recent variant, G.722.2, also known as AMR-WB (Adaptive Multirate Wideband), offers even greater compressions and the ability to adapt quickly to varying compressions if needed due to network changes. Bandwidth is conserved during periods of high network congestion, then returned to a normal level when congestion is alleviated. Because G.722 and its variants sample audio data at a rate of 16kHz, rather than the PSTN PCM method at 8kHz, these codecs result in far better sound quality, clarity, and fidelity.

G.728, or Low-Delay Code Excited Linear Predictive (LC-CELP), coding is widely used in voicemail systems for digitizing stored voice messages. G.729 can deliver an 8-kilobit sample with less than 16ms of processing time. This G.729 codec standard is often used in Voice over Frame Relay (VoFR) and is supported by many frame relay equipment vendors.

The G.729a codec is a variation on CS-ACELP that is rapidly growing in popularity. It compresses the voice stream into 10ms frames. It's less CPU intensive than many of the others, making it a good choice for bandwidth-constrained implementations. It offers a blend of bandwidth and delay performance characteristics that are making it quite popular for VoIP services. It has the added advantage of interoperating with G.729 codecs.

G.723.1, or Algebraic Code-Excited Linear Predictive (ACELP), codecs work differently. They create models of the human voice and “predict” what the next sound will be. These codecs encode the differences between the predicted sound and the actual sound. Only this difference is transmitted to the receiving end, reducing bandwidth requirements. Although bandwidth needs may be reduced, this codec has often resulted in complaints that women and children's voices were not represented accurately.

The codecs described earlier are narrowband codecs, initially developed for supporting narrow voice channels. The PSTN voice channel is a 64kbps circuit carrying audio signals in the range from 300Hz to 3000Hz. Some pure IP voice solutions are now beginning to adopt broadband codecs, mostly proprietary, for VoIP call processing. It's expected that the use of these, and other, new broadband codecs will increase as more and more voice calls are carried over IP in the future.

Voice Call Quality—The PSTN vs. VoIP

As mentioned earlier, the PSTN has been in operation for well over 100 years. During that lifetime, it has been tuned and optimized to deliver one type of traffic, voice, as efficiently as possible with acceptable call quality. There are quality assurances in the traditional PSTN that are sometimes inferred, based on past use, and for enterprise business, sometimes contractually guaranteed. In short, voice service on the PSTN is a known quantity. You know exactly what you expect for call quality when you pick up the telephone.

The PSTN is a “connection oriented” network. When you make a telephone call, the network establishes a voice path, or circuit, across the network that is dedicated to that specific telephone call, and meets all the basic quality requirements you've come to expect. Over the years, the telecommunications carriers have optimized the PSTN to support the delay, jitter, and loss characteristics that guarantee the expected voice call quality. In this connection-oriented, or circuit-switched, network, each circuit is dedicated to a single conversation, so security requirements are easily met. The circuit resources aren't shared as they are in an IP network. Once the call is completed, the circuit is disconnected, or “torn down.”

IP networks such as the Internet share resources all the time. They were designed to carry multiple traffic types. That's the fundamental shift being embraced with convergence. In the legacy past, we designed and optimized a specific network for each type of traffic. A large enterprise might well have a voice network, a data network, and a video network, each with different performance characteristics. Convergence is about leveraging all these technologies—data, voice, and video—onto one single network infrastructure. To succeed, it's important to understand how these different networks behave.

IP is commonly referred to as a “best efforts” protocol. The protocol itself provides routing delivery, but if delivery attempts fail, for any reason, IP will discard the packets. IP was designed to use available network resources in real-time, to deliver data, but without any assurances of delivery and with no quality guarantees. IP was initially designed to support data traffic that is inconsistent and unpredictable in volume, or “bursty.” IP data is described as bursty in nature because different data types have different characteristics. The data flow for email messages or Web browsing is quite different than the data flow for streaming media such as voice or video.

IP is a “connectionless,” or packet-switched, network. Dedicated circuits aren’t established. Messages are divided up and placed into packets. Each packet is somewhat like an envelope in the postal system. It has header fields that contain the source address and a destination address. IP uses routing protocols to identify paths through the network. These packets take the best path available at the time, based on whatever metric a particular routing protocol uses to determine the best path. Thus, as congestion and other factors impact the performance of the network, IP packets can easily take different paths across the network, arriving at different times. A large message is often made up of many IP packets, so they need to be buffered and reassembled at the receiving end.

QoS includes characteristics such as delay, jitter, and packet loss. IP doesn’t provide any assurances for quality or delivery, so performance requirements can only be achieved by adding some other feature, or protocol, to the network. In practical terms, we increase bandwidth, optimize switching and routing paths, and incorporate other, higher-layer protocols to meet VoIP call quality requirements. We also add a layer of security by incorporating industry best practices in network design, incorporating firewalls and user access controls, and including monitoring and response mechanisms throughout the network. Table 5.2 highlights important differences in characteristics of the PSTN when compared with typical IP networks.

Characteristics of the PSTN	Characteristics of IP Networks
Designed and optimized to support voice as a single traffic type	Designed to carry any type of traffic
Uses dedicated circuit for guaranteed delivery	No delivery guarantees using packet-based routing
Connection-oriented—A dedicated circuit path, or connection, is created for each voice conversation	Connectionless in nature—Conversations are digitized and inserted in many packets
Phone calls are typically long duration (4 minutes on average)	IP packets are limited to 65,536 bits; packets are small and only “live” on the network for a very short time; a conversation is made up of many packets
PSTN circuit is dedicated to one 64Kbps voice channel	The IP network uses all the available resources of the network
Delay, jitter, and loss have been designed to meet specific levels for voice calls	Delay, jitter, and loss must be overcome by adding resources to the network
Dedicated circuit provides generally acceptable security for voice calls	Security requires special resources in the network

Table 5.2: Performance characteristics in the PSTN vs. IP networks.

Because IP packets are small in size and data traffic is bursty in nature, IP networks use routing protocols to pick the best path for transmission. In a typical IP network, this bursty nature of data means that the traffic load on the network is constantly changing throughout the day. This constant state of flux means that the best path identified by routing protocols may change often. Routing protocols are dynamic and constantly update to identify the best path through the network based on the current traffic load. The best path one minute may not be available the next. For VoIP calls, this means that the packets comprising a conversation can potentially take many different paths through the network.

Traditional data use of IP is often not particularly sensitive to network delay. If an email message is broken into many packets that can't be delivered immediately, there is likely no problem. The message can be placed in many different packets, then routed across any number of paths available in the network. The entire message can be reassembled and delivered to the recipient, with no repercussions. Unlike email, voice calls are real-time interactions between people. Even a half-second of network delay can be so detrimental to voice quality that VoIP service is rendered unusable.

The PSTN dedicates bandwidth, a 64Kbps voice channel, to each telephone call. It dedicates this bandwidth for the duration of the call. It does this consistently and predictably for every phone call, every time. In an IP network, all the resources of the network are available to every user or node on the network. This is necessary for routing protocols to work and so that IP can make "best efforts" at delivery. Thus, every voice call can be impacted by changing traffic load and network conditions.

Over time, we keep adding new traffic types to IP networks. Today, these networks support email, file transfer, Web surfing, and data sharing. As we add new traffic types to the network, we must redesign and modify the networks to support not just the new traffic type but all traffic types in the aggregate as well. The challenge with implementing and managing a converged VoIP service network is meeting user expectations such as call clarity and fidelity without delay or jitter impairments. Voice conversations must be clear and intelligible or users won't trust the integrity of the network for voice calls. Beyond the quality of the voice itself, the basics of delivering dial tone, successfully completing calls, and always being available are key quality concerns.

Managing call quality perceptions is a bit like walking the razor's edge. Because voice service and telephone calling are so widely used and accepted, user expectations are high, based on their own experiences. When using a new VoIP service, if users feel that call quality is unacceptable, they will simply hang up and use alternative methods. In consumer VoIP solutions, this perception problem is a constant battle. In the enterprise environment, you must take proactive steps during the implementation process to level-set user expectations with what the integrated service can functionally support. If users develop mistrust of the network, even based on misperceptions of capabilities, it is very difficult to regain their confidence. You need to be able to measure, guarantee, and deliver suitable call quality. You also need to ensure that the network can continue to support the pre-existing data traffic without service degradation to existing services.

Network Design Considerations

An often-stated (by consultants, systems integrators, and equipment vendors) misperception is that many corporate networks were poorly designed. It's simply not true. It's important to recognize the truth so that we aren't doomed to repeat the mistakes of the past.

Existing networks in business were not poorly designed, but in many cases, they weren't designed at all. For many organizations, PCs were implemented way back in the past. Windows 95 brought in the concept of the workgroup, and local area networks (LANs) emerged, as small, isolated pockets of information, typically for a workgroup. With routing and internetworking, organizations began interconnecting LANs, and the explosive growth of the Internet was just one more connection.

Today, we have business partner connections, VPN links, Web services, and more. We have the Internet, the intranet, and extranets. We have Web servers, offering Web services, supporting e-business. And in many cases, each new feature or function was brought about by tweaking the service network with some incremental upgrade. We didn't design a large number of our networks, they just happened.

Today's networks have evolved and grown over time. Many have been carefully and methodically redesigned every few years. Like the PSTN has been tuned and optimized for voice traffic, there are IP networks that have been designed and tuned to support known and understood data applications that exist within the enterprise. The enterprise applications in use today may have very different network performance requirements than converged video and/or VoIP services.

VoIP projects begin with an implied assumption that some network upgrades may be necessary. Equipment in the network, such as routers, may already be running at high CPU utilization. In many cases, the router OSs may not be able to support VoIP services. Existing wide area network (WAN) links may already be taxed to their limits supporting existing data applications. In many cases, these links may be using frame relay connections that might not be voice-friendly. Frame relay is very much a data services. The overall capacity in terms of bandwidth and processing power are vital factors to assess before implementing any new network service.

There is a pitfall to watch out for when implementing VoIP services. It's important to avoid the added expense associated with multiple upgrades. Given that for most organizations, network upgrades represent capital expenditures (CAPEX), no technical manager should be put in the awkward position of justifying multiple upgrades. A smart approach is to assess and understand the voice calling and quality requirements in conjunction with all existing data services. In short, invest the effort to do a complete, holistic assessment of the entire network services suite across the enterprise. This is also a good time to consider any planned new data services as well as the VoIP requirements. Rather than upgrading prematurely, focus early in the process on data gathering and information analysis. Taking a methodical approach will not only ensure successful pilot testing, it will help yield a more future-proof design with the ability to grow smoothly and support future, anticipated business needs.

Trunking capacity is always a consideration in voice networks. It's tied directly to carrying capacity. In the VoIP world, how and where gateways are connected to the outside voice network, or PSTN, are important performance and management factors. T1 circuits, whether primary rate (PRI) or standard channelized connections, can provide a standardized approach that traditional telecom providers support. This approach is simply building T1 trunks in the same way you connect an enterprise PBX to the PSTN today. In the traditional T1 environment, a T1 circuit can carry 24 voice calls.

IP trunking presents a different approach, using the full bandwidth of a network connection to the carrier. SIP trunking over IP is becoming very popular and quite common. SIP trunking is easy to implement in most current vendor offerings, and provides an easily quantifiable return on investment. In short, it's proving to be quick and easy to implement, and a powerful cost-savings approach. Given the codecs mentioned earlier, a 100Mb SIP trunk to the carrier, using a codec such as G.729a, can carry a high volume of voice traffic easily.

 Chapter 7 will explore trunking and other services, such as E-911, in more detail in Chapter 7.

The Performance Envelope

There's an old networking paradigm that has been used by many IT managers. When network performance seems sluggish, add bandwidth. This approach works in many cases, but it's usually a temporary fix. It's also an approach that, if followed over time, guarantees that network circuit costs will spiral ever upward. Bandwidth is an important characteristic in network performance, but it's not the single most important factor. Networks have a wide-ranging set of performance characteristics that determine how smoothly they operate.

There's a principle in network management, often related to security, which identifies three important network requirements—confidentiality, integrity, and availability (often called CIA). There are many different components and characteristics that play into achieving complete confidentiality, integrity, and availability in the enterprise network. These characteristics make up what I'll call the network performance envelope. The shape of this envelope gives a view into the “personality” of the network.

We've touched on the importance of knowledge about your network operating environment up to this point. This section will put together the pieces so that you can utilize a knowledge-based approach to ensuring success in deploying integrated VoIP services. The more you know about the network, the more accurate your planning will be, increasing your ability to meet call quality expectations for end users. Better data improves your ability to provide better total network performance.

The basic concept of confidentiality, integrity and availability doesn't offer enough granularity to fully assess the requirements for successfully delivering a converged VoIP service. To accomplish that, you'll need to look at a broader set of data inputs. You'll need specific, measurable data that you can assess to gain a complete understanding of both the existing network and the new requirements.

There are a nearly unlimited number of different data elements an organization might use or consider when defining the network performance envelope. This guide will use the factors shown in Figure 5.1 as an example and will explore some characteristics in more detail than others. These chosen factors provide a good representative sample of the types of characteristics enterprises might consider when working through this process.

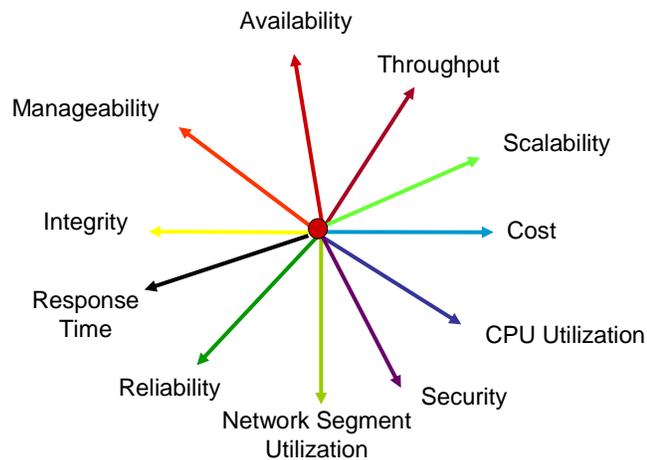


Figure 5.1: Graphing performance characteristics.

This figure shows the basics of CIA in a granular form and brings some other network performance characteristics into the mix to ensure you can fully support converged VoIP services.

Reliability overlaps both the availability and the integrity of the network. Because it's so important, we'll treat reliability as a discrete evaluation factor here. Throughput also plays a dual role, providing both availability and integrity. We'll include three facets of throughput in the example:

- Because bandwidth is a major contributor to throughput, we'll measure it as a standalone factor.
- Response time provides a measurement of delay or latency in the network. As VoIP is an end-to-end service and delay is cumulative, we'll look at response time as a broad indicator.
- CPU utilization may be a performance factor across every element of the network.

We've added cost, which is unrelated to CIA in every way, but a business consideration that must be taken into account with every network change. Some enterprises might substitute a more comprehensive ROI model as part of this assessment. No matter how simple or complex your assessment mechanism, managers will have to answer for profitability and cost recovery of changes to network services.

You must assess the manageability of the network. Deploying new services that cannot be effectively managed is courting disaster and ensures failure. For many service delivery organizations, this correlates into a service level agreement (SLA), or contractual commitment.

Scalability of the network has been included here as a broad measure for future-proofing your design. Every business enterprise plans for business growth. Business growth inherently drives network growth as almost every business sector becomes more and more dependent on data networking resources. You're striving to future-proof the network for some period of time, typically 3 to 5 years. No network designer or engineer wants to implement a major new service such as VoIP on a network, only to discover a year later that the network was under-powered during a major project and now requires a complete redesign.

Integrity encompasses network reliability requirements. Security is included here as an aspect of integrity, encompassing three facets for consideration:

- Packet loss, if minimal, might not adversely impact VoIP services; however, it's a performance component that should be measured and included in your analysis.
- Jitter, or variability in delay, degrades the quality of voice services if not controlled.
- Security can be broken down into many subcomponents. These might include firewalls, intrusion detection solutions, antivirus software, and other tools. For the purposes of this model, security will be treated as a single component.

Availability may have many different meanings to many different people. To some enterprises, and most of the telecommunications carriers, it generally indicates a guaranteed uptime at the five nines, or 99.999 percent level. Rather than deeply explore network design issues such as high availability, redundancy, and business continuity, availability will be used here as another data point on a graph for analyzing the network performance envelope.

Figure 5.2 gives you the next step in the process. It presents the framework you can use for mapping each operational characteristic of the larger network performance envelope.

Network services and applications all have unique and different requirements. Email can be delayed several minutes without any measurable negative impact. Delays in Web browsing might be barely noticeable as the browser screen gradually loads information. Its use of Transmission Control Protocol (TCP) at a higher layer guarantees information delivery, although the overhead might add delay. Web traffic is non real-time traffic, so between a person and a system, performance impact may be negligible. Integrating VoIP into the often already overtaxed data network introduces a whole new set of requirements into the network performance envelope. As these new characteristics needed for new network services are identified, they can be mapped onto the graph. As you do so, notice that the personality of your network, the performance envelope, begins to take shape.

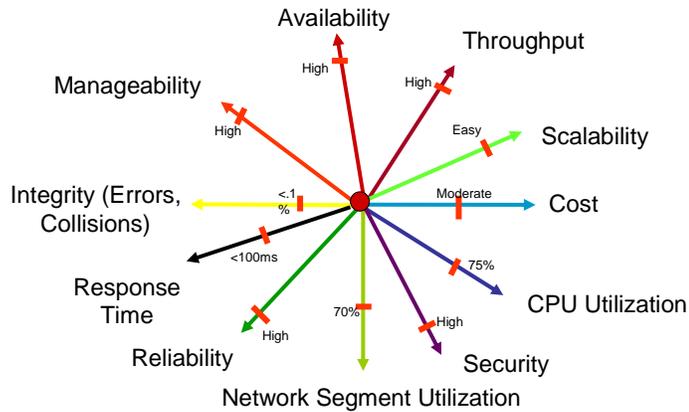


Figure 5.2: Identifying performance characteristic requirements.

Every network service and application in operation needs to be evaluated and factored into the performance envelope data set you're evaluating. Remember, you're identifying the requirements for successful service delivery—your success factors. You are planning to succeed. For existing services such as email and Web traffic, you might make some simple assumptions about acceptable performance. For mainframe applications, or those that were custom-developed in-house, it might be prudent to overlay the specific requirements to support each. The key is that each requirement be identified using some measurable value. The data points or elements used by any organization can be established as appropriate.



This model is demonstrated in a very simple format. Each network has different and unique requirements depending on the services provided and applications in use. If your network employs QoS mechanisms for different traffic types, don't forget to account for each of them when graphing out your network requirements. The more granular the data used, the more accurate the assessment of performance requirements becomes. As always, the more you know about the network, the better your decision will be.

Every performance characteristic you assess becomes a data point on the graph lines. In the next step, which Figure 5.3 shows, “connect the dots” to gain visual representation of the shape, or personality, of your network. You know what the network needs to look like to successfully deliver the existing service and applications in addition to VoIP and other new services needed.

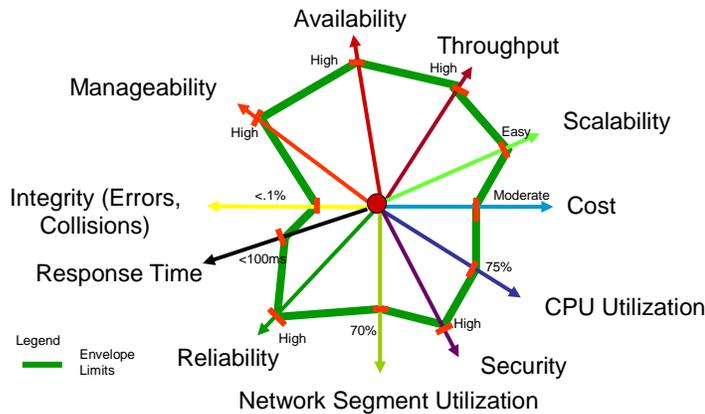


Figure 5.3: Giving shape to the network performance envelope requirements.

For many organizations, now the easy part is completed. The next step in the process is to physically measure each data point on the graph. Some elements, such as cost and security, may be relative assessments rather than technical measurements. The key to success is to be methodical and thorough. This step will quantify network performance capabilities today. As Figure 5.4 shows, when you map the real-world measurements from your network with the performance envelope you've established as a requirement, you have a gap analysis mechanism. This gap analysis can now help you focus directly on those areas of network performance that fail to meet your established requirements. This helps ensure a methodical approach to upgrading the network.

As part of the overall operational requirements of the new, converged network, security and performance management often mesh to become one overarching facet of design consideration. Achieving the delicate balance between service delivery and security requirements means that compromises and tradeoffs will be necessary. The right set of management tools will allow you to continually monitor network performance and security, assess risks, and measure performance around the clock.

Neither performance nor security can be effectively monitored as a single element within the network. The network is a large, almost organic environment. A systematic and holistic view of the health and welfare of the entire system requires vigilance and constant review. Beyond this methodical approach to assessing the network performance, requirements implementing repeatable, sustainable processes will help ensure consistent network performance that delivers both the quality and security necessary for widespread enterprise success. There are a variety of Network Management Systems (NMSs) in use in business networks today. These range from expensive commercial products to freeware and open source tools. Beyond network and security monitoring tools, there are a number of VoIP-specific solutions to aid in constant monitoring of the VoIP service. The best resource for identifying the tools that will work in your environment is generally the VoIP solution provider.

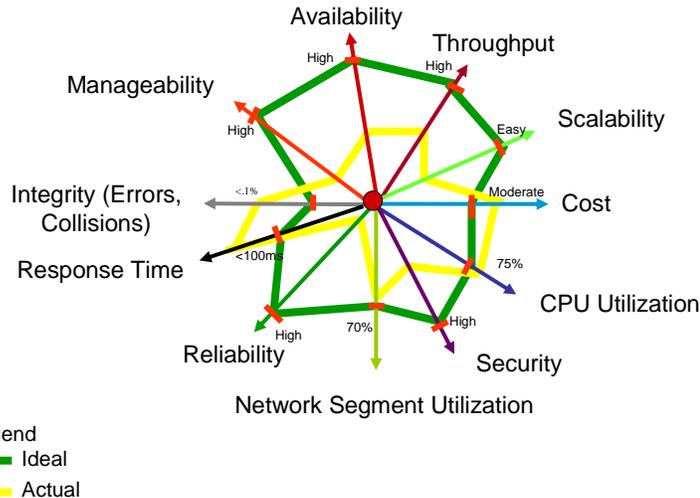


Figure 5.4: Gap analysis—Mapping existing performance against the requirements.

You can use this performance envelope graph to overlay the performance characteristics the network provides today with the requirements you've already documented. You know your requirements for successful service delivery and you know your network capabilities. Figure 5.4 demonstrates a common occurrence in many networks. The reality of network measurements and the requirements don't align in all areas. In some characteristics, the network provides better service than needed. This may mean that you're paying a premium price to a carrier unnecessarily. In other areas, there are gaps to address in order to meet service delivery needs.

Implementing a methodical approach such as the performance envelope when analyzing service requirements also helps in defining VOIP service delivery expectations. Clear expectations are a key factor to a successful implementation. This approach ensures that you know what is required.

Completing a thorough assessment of the existing network has two benefits. First, it provides an accurate and viable gap analysis to use in preparing the network for converged services. In today's frenzied network operations environment, it's sometimes too easy to assume that adding a switch here, a link there, and increasing the bandwidth somewhere else will meet service requirements. It's important to take the time to complete a thorough assessment and provide comprehensive gap analysis information to achieve success. As a side benefit, especially in large, distributed networks, some enterprises will identify areas in which networks have been over-engineered in the past. This can lead to redesign and cost savings.

This performance envelope readiness assessment is an opportunity to re-evaluate the existing network and determine whether current needs are being met. Many corporate enterprise networks were not designed to be what they have become today. Networks often began as small islands of information—isolated workgroups or departmental LANs. As the network gradually grew, connections to other groups and organizations were added. In most organizations, new business applications were also added over time. Corporate networks have grown from a simple beginning into a complex and sophisticated mesh, weaving corporate operations together. Often they've done so without being revisited from a holistic, service delivery perspective. This performance envelope approach to network assessment can help lead to a network that delivers better performance at a lower cost.

Choosing Performance Envelope Characteristics to Measure

Data, voice, and video services each place different demands on the network. They're different *types of service*. Each one is designed to provide a specific service. To support different types of service in the network, you need to be able to offer some consistent and predictable QoS that you can manage. This is often accomplished by creating a specific *class of service* for each type of traffic supported in the network. For some network designers, determining the number of service classes required can be a challenge. Although implementing QoS mechanisms isn't a convergence step that's required, many organizations find that VoIP drives the need in order to deliver the required performance characteristics.

QoS can become very complex. Implementing an overly complicated QoS scheme can lead to a network that can't be readily supported. Rather than add undue complexity, let's look at just some of the performance characteristics included in the performance envelope example.

As network convergence and emerging technologies gain momentum, be mindful of the danger that each new application brings. It's important to look toward the future and not create a new class of service for every new application, or network complexity can spin out of control. For manageability, most network engineers favor using only a few critical service classes.

VoIP and video collaboration represent real-time traffic. This is most often communications information flowing from person to person, rather than interaction with a server or system. These real-time services require quick delivery with quality assurances for delay, jitter, and loss.

IP networks use a best efforts approach to deliver all general traffic. This class of service is quite suitable for email, Web browsing, and most other normal network traffic. Best effort simply uses whatever network resources are available.

In some networks, management traffic may warrant a dedicated class of service all by itself. This approach is quite common in service provider networks. It provides a mechanism to ensure the network can always be managed, regardless of congestion problems that might arise.

All QoS mechanisms provide some form of traffic prioritization scheme. In the converged network, there are many different traffic types. Each may have different requirements and different prioritization needs. Similar traffic types need to be identified so that they can be handled the same way within the network. Most organizations choose to aggregate similar traffic types. This approach allows a company to take advantage of network routes that are optimized to provide the appropriate class of service.

In designing a VoIP service network, the focus is typically on call quality for the voice user, but it's important not to degrade pre-existing services when implementing VoIP. You need to recognize all the different traffic types in use. If mission-critical data applications aren't given the necessary resources through QoS prioritization, the applications might "starve" for lack of resources. Email and Web traffic will still need to be delivered, even if it's a lower priority. It's vital to maintain balance across all traffic types when managing an integrated data, voice, and video network.

Throughput

Throughput is frequently measured in terms of bandwidth. When evaluating throughput requirements, due consideration must be given to traffic aggregation points. Don't overlook the congestion issues that can develop as a result of combining 10Megabit, 100Megabit, and Gigabit Ethernet connections on the network. This traffic, all flowing to a centralized aggregation point, may overwhelm a lower throughput link. This can aggravate network congestion and introduce service delivery problems.

Bandwidth

Over-engineering or over-provisioning the IP network has been a common approach for many network engineers. Increasing bandwidth, by ordering higher capacity links, has been the most common technique. Adding bandwidth may alleviate short-term problems, but it's important to remember that IP uses all the available network resources. Bandwidth is one resource. Over-engineering frequently proves to be a delaying tactic that simply stalls necessary redesign work. This approach can be more costly in the long run. Adding bandwidth still requires investment—upgrading equipment and increasing bandwidth of circuits. These can become very expensive approaches, and they don't solve the problem of design. IP data applications can quickly consume all available bandwidth, leaving the same congestion problem to be addressed.

Response Time

Response time is one measure of network performance, most often measured using *ping* as a test tool. One important nuance in the converged voice and data network is the fact that voice is an end-to-end service between people. Ping can be an effective diagnostic tool, but the end-to-end nature of voice service, coupled with the fact that delay is cumulative, necessitate comprehensive management and testing to ensure service levels and call quality are maintained.

CPU Utilization

CPU utilization in network nodes offers a good indicator of the overall health of the network. In planning VoIP services, utilization might provide an indicator that network elements are overtaxed and can't effectively support VoIP. After VoIP services are live, ongoing monitoring of CPU utilization can provide a benchmark and timeline trend analysis to monitor the health of the network over the complete life cycle.

Network Segment Utilization

Ethernet is, at its roots, a shared media technology. LAN switching provided network technicians a means to segment traffic into smaller broadcast domains. This increased granularity is now often enhanced through the use of virtual LANs (VLANs). Like bandwidth, CPU utilization, and other factors, network segment utilization can be monitored as part of the day-to-day management operations to ensure adequate network performance to support the required services.

Integrity/Reliability

The integrity and reliability of the network encompasses a number of different technical facets. Each can be monitored, measured, and managed as a part of network operations. In the legacy IP network, these may have provided acceptable service and been left unchecked. In the converged service network, they warrant ongoing monitoring and proactive management

Packet Loss

A common measure is error rate and data loss. When IP networks are used to transmit normal data—email, file transfer, Web browsing, and so on—some data loss is acceptable. The higher layer protocols, such as TCP provide a measure of quality assurance and request retransmission when needed. Other data types, mainframe data, as noted, may be very intolerant of packet loss.

Jitter

Jitter describes the variations in delay. As IP networks route traffic over the network using the best path identified by routing protocols, it's possible that every packet in a stream might take a different route. A VoIP call could potentially traverse many network paths. Each route through the network may have different delay characteristics. Jitter typically isn't a concern for the normal IP data traffic. VoIP is far more sensitive to jitter than email or Web browsing because it is a real-time service between people. High jitter can result in unintelligible conversations that sound "jerky." Users won't trust or use VoIP services if the call quality is unacceptable.

Delay/Latency

Delay exists in all IP networks. It exists in several forms. Routers use statistical multiplexing algorithms to process traffic. Assembling data into packets takes time. Checking the routing protocol to identify the best route through the network takes time. These miniscule delays add up and all impact the total end-to-end delay. Delay is cumulative.

Cost

No business can operate without maintaining vigilance in controlling cost or expense. In networking, you face not just CAPEX in hardware investment but also the operating expense (OPEX) of keeping the service up and running on a daily basis. Some businesses, notably service providers, focus on profitability and use ROI as a performance envelope measurement. Enterprise businesses may treat their IT and network operations internally as either a profit center or a cost center. For many, no profit is expected, but the costs associated with implementing and managing the network still must be recovered. For most organizations, over a 3-to-5 year life cycle, network OPEX tends to be much greater than the initial CAPEX investment. It's important to always factor the appropriate cost analysis in both VoIP pre-planning and the ongoing service management.

Availability

Availability, for many, is defined as reliability. It means that the network is fault tolerant and services don't degrade when problems do occur. To provide reliability in the PSTN, there are millions of circuit paths available through hundreds of central offices. In an enterprise IP network, redundant paths, alternative routes, load-balanced connections, and high-availability equipment may all need to be incorporated into the design to ensure resilience in the network.

Uptime—99.999%

Availability in traditional telephony has often been measured as the uptime percentage. In commercial networks, you've heard the term "five nines reliability" (or 99.999 percent uptime) used as the target availability measure. For enterprise operations, it's crucial to recognize that this number equates to roughly 5 minutes of downtime per year. Although the commercial telephone providers have widely met the five nines measure, there aren't many corporate data networks that can claim less than 5 minutes downtime in the past year. Introducing VoIP and other real-time services into the network raises the bar for IP network availability and drives designers to invest in robust, fault-tolerant design solutions.

Security

When implementing VoIP solutions, network security is as great a concern as reliability and call quality. Corporate networks may include firewalls and multiple connection points. These security devices can add nodal processing delay that may impact VoIP services. The more complex a rule set in a firewall, the more latency it induces to the data flow. Remember that delay is cumulative and counts toward the 250ms maximum tolerable end-to-end delay.

VPN services are deployed in two typical fashions. Point-to-point VPN solutions may be used to connect remote offices over the Internet. Many companies also use VPN services allowing employees to connect to network resources while telecommuting or away from the office. Encryption algorithms consume processor power. A VPN device running DES or Triple DES encryption as a VPN endpoint will add further latency as packets are encrypted and decrypted. If the VPN endpoint is a firewall, this CPU load problem may be further compounded.

Security concerns reinforce the importance of assembling a comprehensive technical team for network assessment, readiness testing, and managing the operational environment. It's crucial that the telecommunications, IT, and network security teams collaborate to be successful.

Manageability

The human resources aspect of supporting network services may be the mostly costly component. This is the biggest OPEX cost component. The more complex and difficult the network is to manage, the more resources required. As network designers and service providers, we must consider the ease of management we're incorporating into VoIP network design.

Scalability

Scalability provides the best measure for future-proofing the network. Corporate networks change, grow, and evolve continually. They become very organic in nature as business needs change. The evolutions of the SOA and SaaS on the network are beginning to accelerate. The demands placed on the network will grow as applications, services, and networks become more tightly coupled with business processes. You must always consider the ease with which the network can scale to support new business applications and services.

Summary

The key to successfully managing the converged network begins during the planning process long before implementation. Conducting a needs analysis will help define the business needs in measurable terms. Completing a methodical network readiness assessment will further develop a holistic vision of what the network must support in order to successfully deliver integrated services. Don't sidestep these important steps early in the VoIP deployment process. They will not only ensure that the converged network fulfills your business needs but also help formulate the day-to-day management methods required to support the enterprise integrated services.

Perhaps a key differentiator in using this performance envelope approach to network performance analysis is that it offers a tool to support all services in the network. VoIP is important, but unified communications is only effective when it supports the larger unified application environment that supports the complete business of the enterprise. It offers a holistic view of not just data voice and video but also enterprise-specific applications and services. Sales force automation, inventory management, customer relationship management, and Web-based e-commerce are examples of vital application services that the enterprise needs in order to thrive. Although the software-oriented architecture mindset will more tightly couple these applications with services such as voice and data, using the performance envelope as a holistic approach is an integral part of managing a converged network, encompassing the full spectrum of data, voice, video, applications,

This chapter only briefly touched on the idea of using outside resources. It's worthwhile, in summary, to review how outside integrators and consultants can work for you. They can help look out for your best interests. There are many VoIP consulting firms that specialize in converged technologies. Their reputation is built on how well they take care of client needs. They build business through word of mouth referral and will be your advocate for success. They can ensure that you are being well served by your vendors.

Your equipment vendors and service providers will gladly offer consulting services, and their sales engineers will graciously come in and help you design a solution. It may be pertinent to recognize that vendor staff may be specialists in their company-specific products. They may do a very good job based on their own experience, but they might not have comprehensive training on the workings of competitive solutions. Use your trusted and incumbent vendor partners, but don't give them free rein to design your network. You need to remain in control of the project overall, because you'll have to manage it in day-to-day operations.

The next chapter will delve into the impacts of events in the network, root cause analysis, and how event correlation comes into play as a critical network management tool.

Download Additional eBooks from Realtime Nexus!

Realtime Nexus—The Digital Library provides world-class expert resources that IT professionals depend on to learn about the newest technologies. If you found this eBook to be informative, we encourage you to download more of our industry-leading technology eBooks and video guides at Realtime Nexus. Please visit <http://nexus.realtimepublishers.com>.