Realtime publishers
"Leading the Conversation"

The Definitive Guide™ To

Converged Network Management

Ken Camp

Copyright Statement

Chapter 7: Effective Service Availability Management and Capacity Planning

Life cycle management is crucial to the sustainability of any network. Deploying data voice and video services is all well and good, but long-term success requires a repeatable methodology for consistent measurement and planning. This chapter takes a look at an approach that has been used for many years in the legacy telecommunications industry called FCAPS. The name is an acronym for fault, configuration, accounting, performance, and security management. We look at FCAPS because it has been tightly coupled with managing large voice and data networks for many years. Later, the chapter will touch briefly on the IT Information Library and the ITIL framework. These two well-know methodologies dovetail nicely as a foundation for managing the life cycle of the network.

This chapter will delve into availability management, network optimization, and capacity planning issues as the primary focal point. We'll explore these areas using the FCAPS model as a base for methodical techniques in delivering integrated services.

Network optimization is a vital part of life cycle management. Optimization supports both availability management and capacity planning. It's an ongoing process and aids in keeping network costs in check. Network optimization can also help in managing the demands of business users, who might be rapidly stretching the capabilities or capacity of the network as they deploy new applications. Optimization includes trending, capacity planning, and ongoing minimization of infrastructure costs. By decommissioning legacy services as they are no longer needed and continually monitoring and maximizing the success rate of data voice and video sessions, we are, in a sense, future proofing the network. This holistic approach to network management also provides investment protection by continually evaluating network performance and service delivery throughout the entire life of the network, maximizing the usable life of equipment and leveraging technologies for full value.

Managing the life cycle includes continuous evaluation of both the return on investment (ROI) and the return on effort (ROE). An important part of this process is being able to perform all the following:

- Maximize business potential through performance and availability management
- Learn to decipher results from performance and availability testing
- Use lessons learned from testing to better utilize system resources; this includes assessing capacity and application upgrade needs
- Recognize the risk factors of not testing network applications and capacity, including slowdowns in network response time for your customers or users
- Learn how to monitor, analyze, and prioritize your business and network management needs to ensure a responsive suite of converged data, voice, and video services
- Prepare for the convergence of performance and availability management

These skills often aren't put to use as part of the general IT staff's daily routine. These tasks may actually dilute the focus of the IT group from the daily core business and operational requirements. Balance is crucial so that no single aspect of service delivery, management, or assessment consumes excessive resources. These resources are often viewed as network resources, but the human brainpower and time resources must also be taken into consideration.

Comprehensive network management requires the effective use of people and resources, business processes, technology tools, and products, vendors, and service providers. It's a never-ending effort to provide reliable and consistent service levels for an array of converged services to meet end user and customer expectations.

Introducing FCAPS: A Sustainable Model for Balanced Network Management

As data, voice, and video services converge on the IP network, it's prudent to examine and incorporate the lessons learned by the legacy telecommunications carriers. In many cases, these techniques have also been embraced by the major Internet service providers (ISPs). The telco carriers have been successfully delivering voice services in the Public Switched Telephone Network (PSTN) for more than a hundred years. Although new unified communications technologies such as VoIP, video, and mobile voice are converging to shake up the landscape of telephony, many long-standing service delivery practices are easily adopted to fit today's integrated networks.

The telecommunications industry adopted and followed the Telecommunications Management Network (TMN) framework. This framework was originally defined by the International Telecommunications Union (ITU at http://www.itu.org). Within this large management framework, the general management functionality offered by telecommunications systems is split into five key areas referred to as FCAPS. FCAPS is the International Organization for Standards (ISO) model for network management. It provides a standards-based foundation for network management.

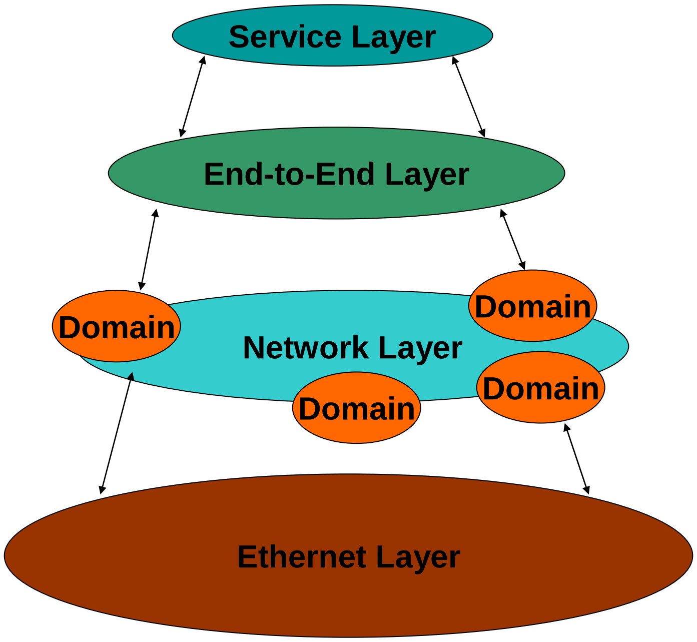

Figure 7.1: The TMN framework.

Digging inside the TMN, the ISO has provided FCAPS as a framework for the task of network management. All forms of management impose elements of monitoring and control. The management of service networks is no exception. For each aspect of network management, there is a life cycle of some sort, corresponding to analysis, design, construction, monitoring, and back to analysis again—endlessly repeating.

There are several areas of mission-critical business management that FCAPS overlooks. The FCAPS model is centered on the technical management of the service network. Daily operational management will include people and staffing issues, finance and purchase issues, process documentation, and a number of other supporting roles. The network manager's day is filled with a variety of tasks and workflows. FCAPS simply provides one network management approach for that aspect of the larger role of a network manager.

Fault Management

Configuration/Change Management

Accounting/Asset Management

Performance Management

Security Management

By isolating the management challenge into distinct areas, the FCAPS model allows us to conceptualize solutions that make the most sense for the challenges unique to each functional area.

Figure 7.2: The OSI FCAPS model.

Fault Management

Every network will encounter faults originating in equipment, cabling, human error, or software during operation. The concept behind fault management is that of formalizing a repeatable process to identify these continually occurring faults. Faults may be recurring events that can be easily recognized with repetition, or they can be discrete events that occur rarely. Smooth network operations require timely isolation and resolution of faults as they occur. A Network Management System, usually deployed inside the Network Operations Center (NOC), provides both trapping and logging of network faults.

Chapter 6 reviewed some of these techniques—using SNMP for trapping and logging system events.

Fault management is not about resolution or restoration of service. Fault management is more a process for intelligence gathering related to network performance. Faults impact performance, and a high fault rate results in impaired or degraded network service delivery. Although mitigations of problems and service restoration must occur in a timely fashion, the fault management flow is oriented to identification and isolation first. Fault resolution and service restore might employ workaround procedures where the root cause of a particular fault may preclude comprehensive correction in a quick and timely manner.

Chapter 8 will explore fault management in greater depth.

Configuration Management

Configuration management is a workflow centered on data gathering and storage. Configuration files and data are collected from network elements. These network elements in the traditional IP network include routers, switches, servers, and so on. In the unified communication network, configuration management is concerned with configurations of gateways to other networks, gatekeepers that authenticate and grant registration access to users, SIP signaling servers, media servers, and even telephone sets.

Knowing where things are deployed in the network and how they are configured is central to managing the total enterprise network. Many day-to-day operational problems are related to improper configurations of one sort or another. Mistyped IP addresses or network masks are a common problem. Patches or updates to OSs, antivirus programs, and business software create gaps in the integrity of the network. You need to monitor and manage all aspects of configuration, including:

- IP addresses
- Subnet masks
- DNS settings
- Frame size
- Directories
- Disk drives
- Drivers for network cards
- Video cards
- User groups and identities
- Domain structures
- Applications

📖 The list goes on and on. Chapter 8 will explore configuration management in more detail.

Accounting/Administration/Asset Management

Comprehensive accounting and effective control can save money and improve the allocation of scarce network resources. How do you know whether an upgrade is needed without measuring usage? Factors that might be measured include printer pages, use of disk space, processor time, network bandwidth, use of resources (for example, the Internet), and so on. Choosing an appropriate metric is often tricky. For instance, do you measure the number of pages or the amount of ink used for printer accounting? The latter is where the cost is.

Accounting costs money and should be used only where appropriate. Costs can include computer resources (such as disk space, processor time, network bandwidth) as well as the time and effort spent enforcing restrictions and collecting funds. A large portion of the cost in your legacy phone bill is taken up by the cost of collecting it.

The FCAPS model evolved from the legacy commercial telecommunications industry. Customer billing for minutes of use is a core component of that industry segment. In the FCAPS model, early documentation focused on these traditional customer billing, or call accounting, systems. Efforts to mirror this accounting mindset in data networks have been implemented through tools that collect usage statistics that may be tied to disk space, CPU utilization, minutes of use, bandwidth, or some other metric. Bandwidth has become a simple metric that's widely used. Tools implementing protocols such as RADIUS and TACACS are commonly deployed in enterprise networks to monitor usage. For most enterprises, auditing, rather than accounting and billing, is the primary usage for these monitoring systems.

In enterprise networks whose core business is something other than telecommunications or voice service delivery, *administration*, *asset management*, or *auditing* may provide a more accurate frame of reference for this piece of the puzzle. Administration objectives include managing a set of authorized users by establishing user IDs, passwords, and permissions. Administration can also include support of systems and equipment such as performing software backups and synchronizations between systems or network elements.

Asset management might simply encompass the financial aspects of capital investment and system replacement as asset costs are amortized off the corporate books. Beyond physical hardware assets, many enterprises adopt their asset management efforts to include software licenses and oversight of version control.

Auditing, in the enterprise world, can range from simple usage auditing to enterprise policy management. With the widespread focus on regulatory and compliance requirements—such as the Sarbanes-Oxley Act (SOX), the Graham-Leach-Bliley Act (GLBA), and the Health Insurance Portability and Accountability Act (HIPAA)—audit processes often expand. Business sectors may have other auditable compliance requirements based on business sector.

Performance Management

Performance management is another aspect of the network that looks solely at performance data. Continual monitoring of network health and performance requires collection of performance data. Although SNMP might be used for monitoring faults and errors as part of fault management, performance management monitors interfaces on network elements, throughput, responsiveness, and the holistic performance health. In a converged network that delivers unified communications services, performance management drives the monitoring of trends in network performance and utilization. These trends can raise the visibility of network capacity and reliability issues before they impact delivery of the suite of unified communications services. Performance thresholds can be set to trigger an alarm, which might then be handled through the fault management process.

To monitor and improve the service, deploy hardware and software effectively, and to spot trends and changes, monitor and control performance. Closely allied to fault management is the task of measuring everything from the number of bits sent per second to the speed of transactions to the refresh rate of your monitor. Understanding the things that can affect performance is a technical issue, but knowing what to monitor and why is a management function. You need to establish baselines and identify bottlenecks in the system. Bottlenecks are the slowest point in a system. A lot of money can be wasted attempting to improve a system if you do not know the part that is making it slow.

> 🕮 See Chapter 5 for more detail about performance management.

Security Management

The main objective of security management is to identify and mitigate risks. In the service network, security helps:

- Maintain legitimate use of corporate resources
- Maintain confidentiality of enterprise intellectual capital
- Ensure integrity of the data integrity
- Provide consistent auditability

Security is a big issue and always comes with attached costs. These costs can involve extra processing (for example, for encryption, authentication), extra administration (for example, setting up user IDs, monitoring for security breaches and so on), extra hardware (for example, firewalls and fiber optic cable instead of easily tapped wires), and a host of other less tangible elements, such as convenience or lack thereof, that might impact the ability of users to work effectively.

> 🕮 Chapter 9 is devoted to network security risk and incident management.

FCAPS Simplified

Today in business network management, this FCAPS model exists in many forms and flavors. As defined by the ITU, it is more complex than necessary for many organizations. It's important to remember that FCAPS was designed to support the telco carrier networks. It may be the best choice for a VoIP service provider but not for a health care or financial institution. Rather than undertake the complexity of a full-blown FCAPS methodology, many organizations implement a four-layer approach for converged network service assurance focusing on:

- Performance management/availability management—Guarantee availability and performance of the network infrastructure, VoIP, and other application or business services
- Security management—Ensure protection from risks and response to security-related incidents
- Configuration and vulnerability management—Manage network configurations to ensure compliance with corporate guidelines; this could be an enforcement point to ensure regulatory compliance where applicable
- Change control management—Ensure that only authorized and appropriate changes are made to the network; audit controls provide knowledge-based tracking of every change

You've often heard the acronym TCO. Throughout, this guide touches on four facets of holistic management for the converged services network. These are:

- Operational integrity
- Service management
- Policy compliance
- Risk management

	Performance Management—Availability Management	**Security Management**	**Configuration and Vulnerability Management**	**Change Control Management**
Operational Integrity	Monitor system and network performance	Continually monitor threat environment	Identify vulnerable or exploited systems	Ensure no unauthorized changes are made to operating environment
Service Management	Map performance indicators to business issues	Event correlation of security incidents with performance issues	Benchmark and baseline system configurations	Delegate admin change permission only to approved staff
Policy Compliance	Measure SLAs and compliance	Protect audit trails, monitor and report security violations	Ensure baselines follow corporate standards and policies	Audit logs for all changes to production environment
Risk Management	Forecast and avoid service interruptions	Identify, mitigate, and respond to security incidents	Identify, assess, document, and remediate or accept risks	Test changes before implementation to minimize risk

Table 7.1: A four-layer variation on FCAPS.

The various capabilities identified in Table 7.1 can all be implemented at once. Many companies today follow the ITIL framework as a root model. The ITIL framework is a series of industry best practices. It’s a set of documents that focus on the core business areas of IT service management (ITSM). ITIL is referred to as a library because it has been published in a book series. ITIL and IT Infrastructure Library are registered trademarks of the UK Office of Government Commerce (OGC), where this concept took root. The library has evolved to encompass a wide set of best practices.

When you think about adopting best practices, there is one simple view that makes their benefit clear. Best practices provide a practical substitute for conducting comprehensive risk assessment and analysis. In most cases, the global business universe can leverage the shared knowledge gained through others and lower the cost of progress by avoiding the high cost and effort associated with comprehensive risk analysis best practices.

The ITIL framework supports the wide field of IT infrastructure, development, and infrastructure for service delivery. One facet of ITIL is referred to as the *maturity model*. When organizations embrace ITIL, they typically conduct a series of assessments of their own process maturity. Most companies choose to work methodically toward a mature model that supports all the enterprise business needs. They adapt and refine along an evolutionary path and adopt new methods and techniques along the way.

Optimization for Service Availability and Capacity

Approaching at service management, availability, and capacity planning from the traditional carrier telco provider's perspective wasn't necessarily easy to adopt early on in the evolution towards the converged networks of today. For most businesses that were early adopters of new technologies such as VoIP, it seemed much simpler to add to the existing network and grow incrementally. This controlled growth can be a good thing from an investment and impact perspective. The danger is in cobbling together disparate pieces and continuing to add and grow without stepping back to take a larger view of what is happening in the network. The performance envelope discussion in Chapter 5 provides one technique for protecting against growth pains through methodical evaluation.

The larger the enterprise organization, the closer the proximity to the carrier's mindset. Many large enterprises, especially in the Fortune 1000, find that over time they become their own telephone company. They have become their own data network provider. These organizations provide internal voice and data services to more employees than many small telcos and ISPs do. And they add the complexity of business processes to support their core business, which might be financial, health care, transportation, or some other facet of business.

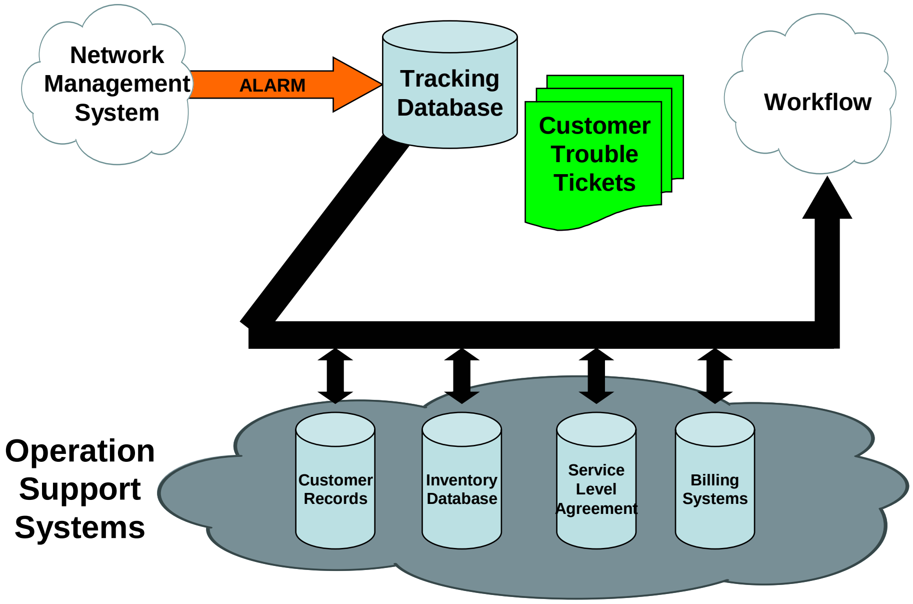

Figure 7.3: The typical tools.

Figure 7.3 shows the typical tools of the voice and data network service delivery business. Network management systems monitor the service network, producing logs and alarms that feed into databases. Some form of trouble resolution mechanism or Help desk tracking is needed. Operational Support Systems (OSSs) provide repositories for other information—customer records, inventory, contracts or SLAs, and billing. The cloud in the visual of workflow represents all the rest of the core business and includes the processes and procedures around service delivery in the corporate environment.

The key is in scaling the methods described here to the size and scope of your business. Customer records alone can vary from a major Customer Relationship Management (CRM) system, running on widely distributed servers, to a central contact management system in a mid-sized organization to an individual salesperson's customer database running on their desktop. For small businesses, the contact list in the salesperson's cell phone might be the customer database. At some level, every component the carriers deploy to manage a global telecommunications network correlate to similar functionality in every enterprise organization.

Trending and Capacity Planning

This guide is really aimed at service delivery organizations, whether they are a VoIP service provider, the IT group in a small to mid-sized business, or the Chief Information Officer (CIO) in a Fortune-100 corporation. Every business depends on business intelligence for survival and success. A technology organization inside a large enterprise that is in the business of delivering services to that enterprise has just as great a need for business intelligence as a provider of services in the open market.

Monitoring trends and usage provide the supporting data for ongoing network expansion and upgrade. Although a NOC might focus on monitoring for availability, uptime and performance, the capacity planning staff needs comparable data. It's not enough to manage utilization of trunks and network connections. Capacity planners need to know what features are being used too. Feature sets such as call hunt groups, call pickup groups, and automatic call distribution groups may be tied to levels of licensing. Given that many VoIP systems are now servers running on general-purpose hardware platforms, software licensing and upgrade has to be considered in the capacity planning process.

Bandwidth

Bandwidth has become, for many, the holy grail of networks. The consumer end of the spectrum has evolved from very slow dial-up connections to broadband in the home. The enterprise network has gone from fractional T-1 circuits over frame relay to 100Mbps switched Ethernet to the desktop and Gigabit Ethernet in the core. Large enterprises may even use SONET technologies to provide even greater bandwidth across a wide area network (WAN). Dark fiber deployment during construction and around the country continues to increase. Wireless technologies are spreading throughout municipalities and are quickly becoming prevalent in campus, warehouse, and office deployments.

In the PSTN and Internet, high-bandwidth circuits have carried the backbone traffic for years. Unified communications technologies have, over time, pulled the two clouds closer and closer to the point that today lines of distinction between the two blur. This is a challenge for the telecommunications and Internet providers, as they've often used bandwidth and transport technology as a differentiator. Customers and end users care less about technology and more about solutions today. As described earlier, what matters is that sales force automation or CRM systems, and others, simply integrate and work.

Costs of fiber, higher-speed Ethernet switches and routers and advances in optical technology have driven the cost of higher-capacity connections down to some of the lowest *dollars-per-bit* prices ever imagined. In terms of VoIP deployments in particular, this low cost can induce motivation for change and is something that needs to be evaluated, designed for, and monitored for the life cycle of the network service. In the past, enterprises deployed Private Branch eXchange (PBX) systems for their internal corporate voice services. These systems were typically connected to the PSTN via T-1 circuits carrying 24 voice channels. An ISDN Primary Rate T-1 was the most common PBX connection.

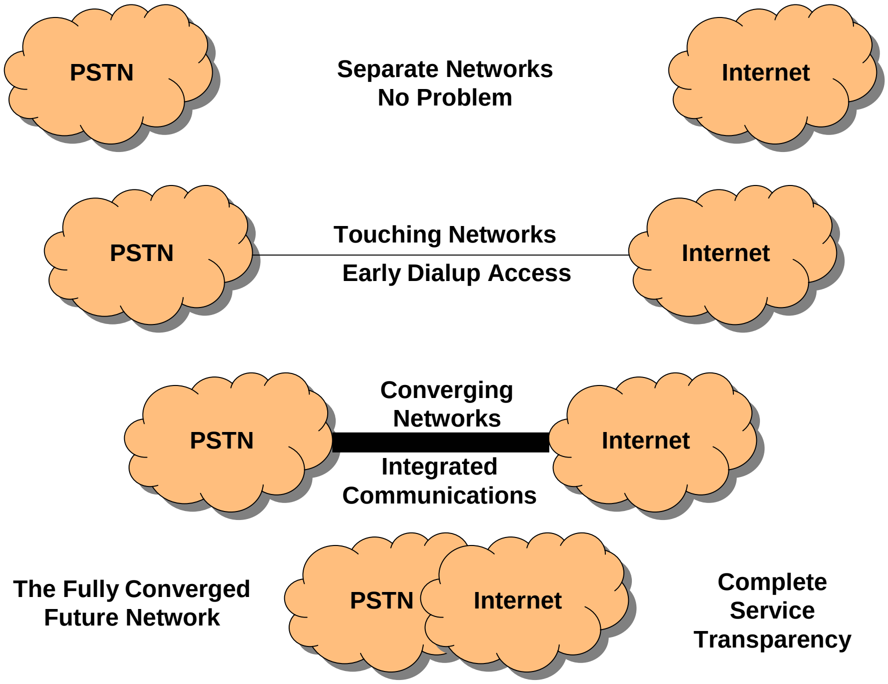

Figure 7.4: Network convergence evolution.

This is a good point to revisit the network convergence evolution from Chapter 1. Over time, completely separate networks, the PSTN and the Internet, have evolved to become tightly coupled. Today, they complement one another to deliver an array of data, voice, and video services.

This tight coupling of the PSTN and Internet has led to choices in design for service delivery today. When providing voice services, one decision will be selecting the appropriate trunking technology between networks. For enterprises that continue to maintain a traditional PBX, T-1 trunks to the PSTN might represent the best solution. Some organizations will implement gateways that convert from the IP network in the enterprise to the PSTN network via T-1 circuits.

As enterprise networks move forward to integrate communications services, SIP is quickly gaining momentum as a peering and trunking technology between voice service networks. SIP trunking may introduce some new costs, such as Session Border Controllers (SBCs). These provide a gateway and firewall services between service networks. SIP trunking is rapidly proving to be very cost effective.

Bandwidth comes into play in SIP trunking because they're generally IP network connections. If the cost of a 10Mbps or even a 100Mbps IP link using Ethernet is significantly lower than a T1, many more voice calls can be carried at a fraction of the cost between connected networks. A legacy voice T-1 circuit carried 24 phone calls. IP links can carry many, many more.

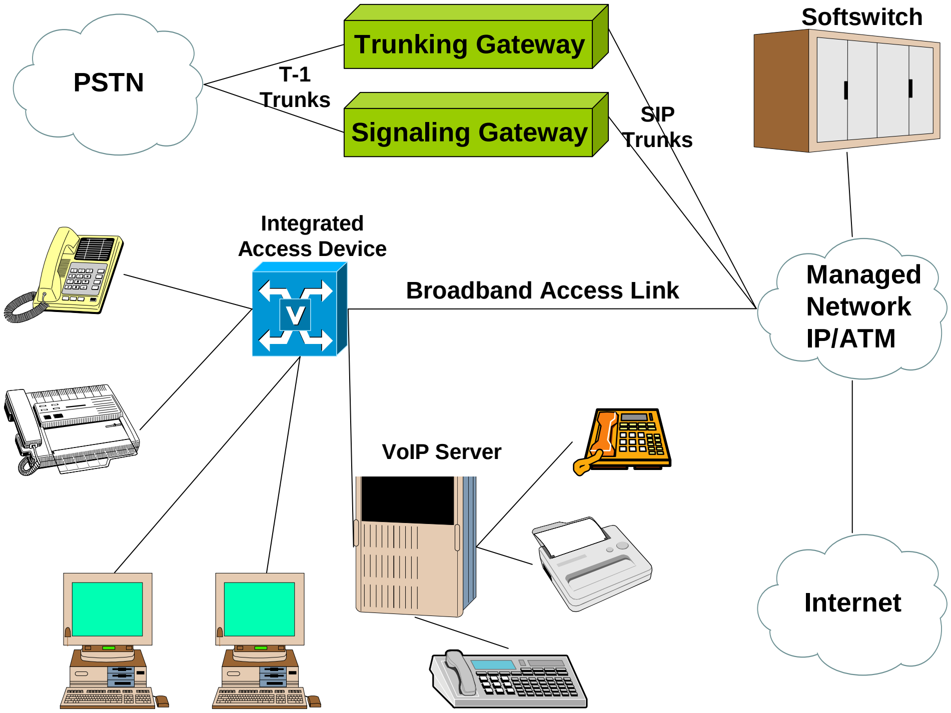

Figure 7.5: The variety of convergence options.

A glance at Figure 7.5 reveals the variety of connection times coming into play as technologies converge while they're evolving. In some cases, separate signaling and trunking gateways are apparent. In other cases, they'll be integrated into a single softswitch.

SIP trunking may be used between what was thought of in the data environment as extranet business partners. For example, a health care provider might implement SIP trunks to connect o insurance carriers and pharmaceutical providers as integral partners. SIP trunking is also on the rise as a connection to both traditional telecommunications carriers and other voice and video service providers. Like every other key component of the converged service network, bandwidth must be monitored and managed to ensure service levels can be maintained.

Ports/Lines

Ports and lines are a primary consideration during the initial design phase of any VoIP network. The number of trunk ports and telephone sets a system can support is rarely overlooked. Capacity planning needs might seem simple, but can be deceptive.

The service delivery team must account for business plans as well as technical plans. Mergers and acquisitions can radically alter the number of trunk ports or telephone sets a company needs. It's vital that the technology service delivery organization participate in business planning, or at least have clear vision of where the enterprise is moving in order to support the wide range of business needs.

Codec Planning

Chapter 5 explored codecs, primarily for their role in network performance. Codecs directly impact the fidelity of the voice that the end user hears. Codecs have other impacts as well. Remember that you hear an analog voice sound. The process of compressing, digitizing, and packetizing voice requires hardware. The choice of codec can drive CPU utilization up or down.

Codec selection also impacts the bandwidth requirements inside the service network. Pulse Code Modulation (PCM) was designed for the 64Kbps voice channel of the T-1 architecture of the PSTN. However, the G.729a (CS-ACELP) codec that has been a rising star in VoIP solutions recently can be optimized to produce a 6 to 8Kpbs bit rate. Although the real world doesn't generally produce a true 8-to-1 reduction in bandwidth required, this codec can dramatically impact bandwidth requirements.

It's a good idea to regularly re-evaluate codecs in use. One practice that is becoming more widespread is an annual codec review with an eye to quality of service for end users. Although converting to a different codec seems labor intensive, it may be one approach to improving quality if the network is becoming saturated.

Optimizing the Infrastructure Optimization

Chapter 5 looked in depth at the performance envelope approach in the context of network readiness assessment. Just as the life cycle is continuous, so is the evaluation or assessment process.

Optimizing Hardware

In legacy PBX technology, the life cycle of an investment was commonly projected to be 10 years. In IP networking, the refresh cycle has become much shorter.

Repair is most commonly viewed as trouble resolution. For PBX technology, it often means swapping out parts. The same might remain true for a VoIP call processing system today, especially for mechanical component failure. Disk drives and fans will fail. Power supplies will burn out. Hardware will require repair.

Another perspective worth considering is that the integration of unified communications services moves off traditional, dedicated, and often proprietary hardware to what might be viewed as general-purpose systems. Most enterprises standardize on a server hardware platform that often supports VoIP services as readily as Web services.

You need to not only consider the legacy telephone system at the time you initially deploy VoIP services. Because voice and data are becoming integrated, the data network life cycle also comes into play. For some organizations, obsolescence is only one driver. There may also be compelling business reasons to migrate with newer technology. As discussed earlier in this guide, the ongoing integration of software applications and network service can make it prudent to move to new technology—not when the older systems are at the end of life but when the end is visible on the horizon.

Another important consideration is that manufacturers discontinue support for older products every day. Some do so because parts become scarce, making it more difficult to repair systems. In many cases, vendors declare “end of life” of a product hoping that customers will migrate to newer systems.

As systems age, maintenance and support tend to become more expensive. Newer hardware brings new efficiencies in patching and upgrading, which may actually drive support costs down. And older systems may just reach full capacity with no expansion capability to meet current and future business needs. When evaluating the refresh cycle for systems approaching obsolescence, it’s wise to remember that technology won’t do you any good if it can no longer support your business needs.

Optimizing Software

The ongoing management process must also encompass software. In the old world of PBXs, software was often proprietary. An upgrade or patch was tested and businesses deployed it into the production phone system at night or over a weekend when the phones weren’t highly used.

VoIP service elements are often general-purpose hardware; maybe even the same platform as the company’s data servers. Software optimization may include modifying configurations, adjusting frame or packet size, or even some esoteric nuance like altering the TCP window size in the registry setting

Maintenance, Enhancements, Upgrades, and Patches

Many companies understand the need to keep the service processing software up to date but overlook underlying operating systems (OS). Whether a voice service element is running an OS from the UNIX/Linux family, the Windows family, or a vendor's custom internal OS, it's important to incorporate a patch management process that ensures timely updates of both the OS and the service software. For a large service network, this will often necessitate a lab environment for testing all upgrades before they are deployed in the production network.

SLA Optimization

Companies that contract with external providers watch SLAs closely to ensure contractual obligations are being met. In the enterprise network, SLAs might only be described in service offerings and not actual contracts to business units. Internal SLAs may be quite informal.

In the enterprise world, VoIP service is often designed to meet specific criteria for successful internal delivery. Characteristics for the quality of service needs—such as delay, jitter, loss, and other QoS provisions—need to be handled just like contractual SLA commitments. As conditions change, the network is regularly tuned to meet these SLA requirements, whether they're contractual or self-imposed to assure the service quality.

It's also good to remember that the SLA approach isn't just documentation of the minimum service levels being provided. It's also used to describe remedial actions to be taken. For internal service delivery, the SLA can be a process document with specific details for life cycle management. It may also be worthwhile to document what service classes, or bandwidth and QoS commitments, are being delivered to end users, then monitor for over-usage. In the enterprise network, it's common for internal customers (in business units or divisions, for example) to evolve and suddenly consume far more network resources than originally planned.

Although enterprises outside the service delivery sector may have no need for the SLA-driven model, there may be other motivators for the larger enterprise. When SLA-based network monitoring and management systems are fully integrated and automated, they can provide both long- and short-term visibility to overall and service-specific operations. Adopting this "service provider mentality" turns the SLA into a key tool for successful operation. The cycle of continuous improvement driven by managing the SLA metrics can, and should, be modified over time as network objectives and realities change and evolve.

Delay

Delay is simply a reality of IP networking. Delay is cumulative and all sorts of factors can impact the time it takes for an IP packet to traverse the network. In addition to delay due to transmission distance, remember that there are other processing delays. In most enterprise networks, delay is usually a constant factor for any given communication. Network traffic in most enterprises tends to follow the same path for everything of that media stream type.

Call Setup

Delay in VoIP networks can lead to call setup failure. For most implementations, this is a factor that doesn't have an adverse impact. Rather, this is a consideration for ongoing management and monitoring of the VoIP service network.

During Call

Regular monitoring of delay in the network and knowing the operating environment may lead to quick trouble resolution later. If you recognize that data traffic is *bursty* by nature and that traffic such as email and Web browsing are not real-time traffic, it's easy to see that they might not provide good early indicator of rising delay during day-to-day operations.

VoIP users may provide the "canary in the cage" sort of early warning system that helps in managing the service as a whole. A change in delay patterns in the network, coupled with routine monitoring and a few calls from users identifying call problems or call setup issues, can quickly help identify problems and lead to mitigation.

Delay Variation

Delay variation is called jitter. Because packets can take different routes across the network, delay variation in large IP networks is common. Jitter is measured in milliseconds. For most enterprise networks, no extra engineering is needed to maintain a 1 to 2 millisecond level of jitter. Large enterprise and service provider networks, spanning across a large geographic area, are more likely to have diverse routes of varying distance that can increase the tendency for jitter.

Jitter is generally not a problem during call setup. It's more likely to arise during the media flow associated with a conversation. In other words, jitter is more likely to impact the voice than it is the signaling. Jitter in voice conversations results in unintelligible conversations that sound "jerky." Jitter buffers can be deployed to reduce this impact, but buffering can only do so much to overcome impairment.

Jitter is managed in a number of ways. Many vendors' equipment provides buffer management algorithms for setting the size of ingress and egress buffers. Beyond buffering at the node level, some vendors include jitter buffers. These are very small, often only 3-bits in size. They can be used to compensate for the timing variations. These jitter buffers may be static, dynamic, or adaptive. Static buffers are common in older routers, particularly in small branch office or home office devices. Dynamic buffers calculate an optimum buffer size based on the first series of packets received. The most advanced jitter buffers adapt to changing network conditions. There are unified communications monitoring tools in both commercial and open source variations that enterprises can use in the NOC to include jitter monitoring as part of the overall network health watch.

Packet Loss

Remember that IP is a best-efforts protocol. It relies on higher-layer protocols, like Transmission Control Protocol (TCP), for delivery guarantees. Packet loss will occur. It's also referred to as error rate.

Blocking/Non-Blocking Access

In the traditional telecommunications history, engineers design central office and PBX systems to be either blocking or non-blocking. A non-blocking system assumes that every end user will be able to go off-hook at the telephone set and place a call at the same time. In the traditional environment, non-blocking systems were very rare. Building a non-blocking system meant investing in hardware and circuit connections that statistically would sit idle most of the time.

With the evolution to VoIP, and the shrinking cost of networking, some organizations take a different view today. Non-blocking systems may be achieved at a lower cost, and may be the appropriate solution for some businesses. If a non-blocking system is put in place, the monitoring systems may need to evolve to track items such as successful versus failed calls and abandoned calls to assure that a non-blocking service delivery network is being maintained.

Success Rates of Call Setup

During network readiness assessments, engineers evaluate current requirements of the data network and compile existing voice services requirements. One aspect of ongoing oversight in the voice services network is close monitoring of the call completion rate. Calls may fail to complete for a number of reasons, but a low or declining call completion rate might be an indicator of signaling problems, bandwidth consumption (or a network saturation problem), or a number of other issues.

Another aspect of call completion rate to monitor is the number of abandoned calls. In a call center environment, abandoned calls are generally tied to wait time for customer service agents. A message telling a caller that the wait is 10 minutes may induce callers to hang up and retry later. Interactive Voice Response (IVR) systems or dial-prompt systems with deeply nested menus may drive callers to hang up in frustration. Call completion rate trends can provide information about specific VoIP services that make capacity planning and overall service more effectively support business processes.

Gateway Issues

Because you're looking at a converged network of voice and data, you have to recognize that not all organizations are going to move directly to new VoIP systems. Convergence is a drive to integrate existing services that work together in new ways, then managing the new network environment of voice and data.

Figure 7.6 shows a company with old and new. The location on the left has a PBX connected to the VoIP gateway. In this case, the PBX programming determines which calls are directed to the PSTN and which are directed to the VoIP gateway. This might be done via a special access code, or it might be programmed into the dialing patterns of the PBX. For example, any time a 10-digit number with area code is dialed, the call might be directed to the IP network to minimize long-distance charges on the PSTN. The company might route only local and 800 number calls directly to the PSTN.

The gateway processes and packetizes the phone call, then passes it on to the router, which will transfer it along to the corporate IP network. At the other end of the network, on the right side of the figure, notice there is no PBX. This company has chosen to eliminate the cost associated with buying, managing, and maintaining a PBX at the location on the right. There is a mix of VoIP phones and softphones at this office.

The SBC on the right is using SIP trunking via the IP network. VoIP calls come into the SBC, and are then routed through the network to the appropriate endpoint. When everything is configured properly, every voice endpoint can talk with every other endpoint. Traditional voice and VoIP have converged.

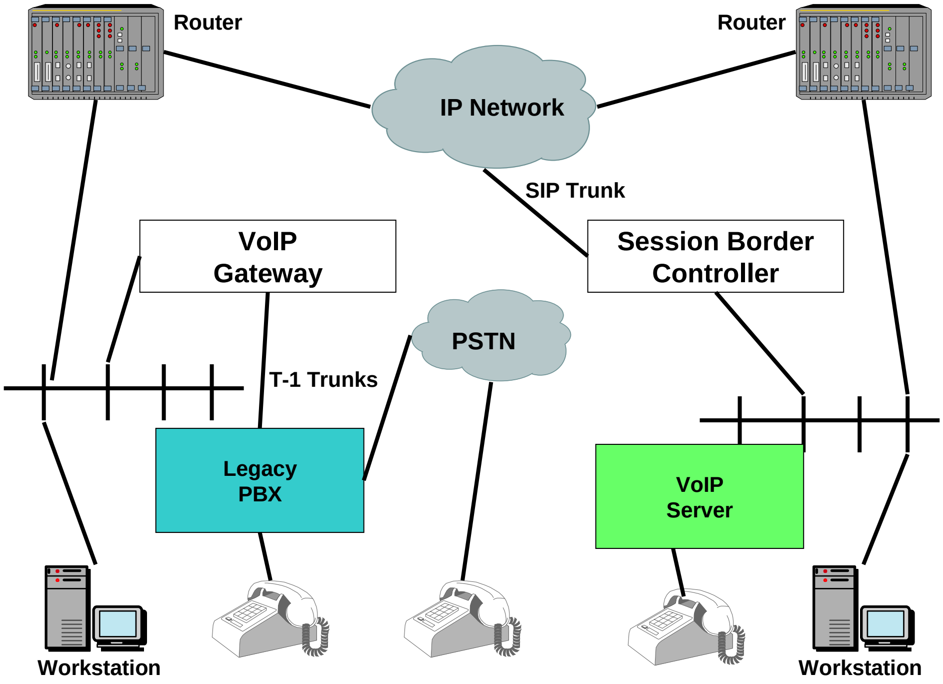

Figure 7.6: A converged company example setup.

This highly simplified view of the network might make convergence sound like a trivial matter. It is not. This company is managing a complex multi-faceted network, and performance monitoring will have to be undertaken diligently.

The gateway must have a database of telephone numbers associated with users, yet you must not overlook that many LANs use Dynamic Host Configuration Protocol (DHCP) for address assignment at startup. Thus, the gateway must maintain a dynamic database that allows a static telephone number to correlate to a dynamic IP address every time the user connects. The SBC simply represents another type of gateway where SIP trunking information and access permission controls reside.

Monitoring systems need to provide alarms for the NOC staff to alert when a number of new situations arise.

- DHCP addressing is straightforward, but how do you monitor and alarm error conditions when an unknown user tries to register with the VoIP system?
- Trunk utilization between traditional and VoIP systems, whether done in house or with a telco provider, must be monitored closely. Trunking or media gateways often prove to be the choke point in voice call services.
- SBCs require the same sort of monitoring as a major enterprise router or firewall. They represent a new perimeter point in the network.

Although convergence presents many wonderful opportunities to bring services together onto a single infrastructure, it also brings several added degrees of complexity to the network. The existing IT staff may not have a deep enough understanding of telephony requirements. The telecom staff may not be familiar with data networking technologies. The skill set required to design and manage this network is a blended skill set that has not been necessary in the past when the two networks remained distinctly separate. As networks and services converge, the pool of talent with the right skills to design, manage, and maintain those networks shrinks. It's important to ensure the support staff receives the proper tools and training to keep these mission-critical services operating at acceptable performance levels.

Network management is one of the critical and most challenging aspects of service delivery. In the past, many network designers followed the simple rule of "better safe than sorry." This often led to over-engineered and over-priced networks.

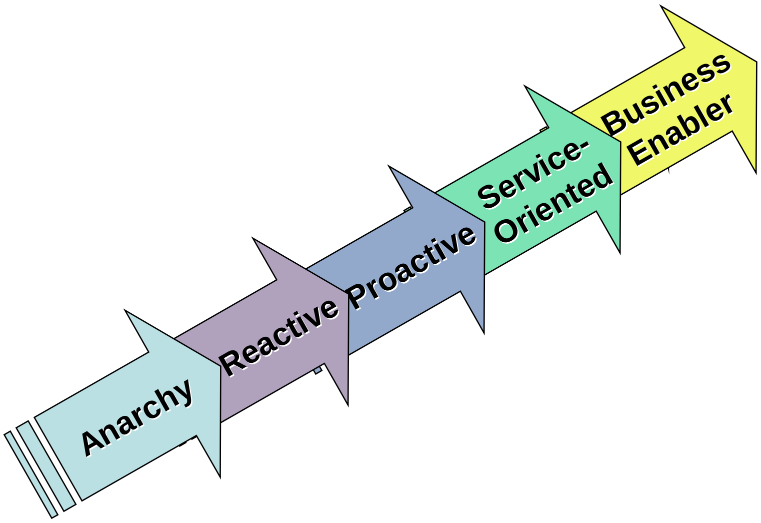

Figure 7.7: The network service maturity path.

Figure 7.7 shows a network services evolutionary path that is actually quite common in business organizations. The underlying goal for focusing on managing the integrated data, voice, and video services is to turn them into business enablers for the enterprise. This evolution of maturity follows a parallel track with the ITIL maturity mentioned earlier.

If there is a complete lack of business process, anarchy and chaos reign. No business manager plans this, but sometimes explosive growth, especially in a highly successful start-up company, can cause total chaos. Early business process development is often reactive in nature. As business processes mature, you find they move into a more proactive, forward-looking model.

For the most successful information-oriented enterprise, network management becomes a part of the service. It becomes a differentiator. And in the most advanced cases, network management information is such powerful business intelligence that it is leveraged to enable completely new lines of business.

Summary

Knowledge about your network is a crucial element to effective service delivery. Service delivery managers need know and understand every aspect of network service delivery. Without a knowledge-based framework, service assurance efforts are hit or miss at best. Managers overseeing services networks need to know as much as possible about performance trends and service requirements to guarantee availability of mission-critical services in the enterprise. As you integrate voice, through VoIP, with the IP data network, you migrate what is for many organizations the single most mission-critical service. Capacity planners also need utilization information to provide consistent, systematic growth across the suite of unified communications services as business requirements evolve. Everyone involved in service delivery needs information about the user experience to identify bottlenecks in service availability and delivery. Correlating information about failure points is vital to effective troubleshooting in daily operations.

Industry best practice today, whether via FCAPS, ITIL, or some other model, all encourage taking a holistic approach to network life cycle management. That holistic view is widely supported by a wide-ranging data gathering and information analysis strategy. To ensure service delivery, you must align service delivery efforts with the needs of customers, internal or external, while finding a cost-effective approach to ensuring capacity, availability, and security of the data, VoIP, and video services.

Two of the most treasured assets any information-based company has are its people and its intellectual capital. Data, information, and knowledge are frequently far more precious than inventory and cash reserves. The latter are generally easier to replace or regenerate. Raw data is easily collected. In years past, log data often either purged from systems after some time period or went into archival storage. For many organizations, historical information that you now perceive as key business intelligence was stored on tape or microfiche and buried in a data tomb. In more recent years, attentions have focused on data warehousing concepts. This quickly spawned another industry subset promoting knowledge management solutions. Today, many of these knowledge management theories and techniques are used to arm managers with information needed to ensure service levels meet user needs.

The enterprise service delivery workforce requires appropriate data gathering and analysis tools to facilitate making informed decisions about network services in daily operations. The more accessible and readily available these tools are, the more efficient and productive the organization can be. One way to bring the tools closer to hand is to create that converged environment where access to any piece of information and any communication need is available instantly. This knowledge-based approach ensures the highest level of confidence in all network services.

A well-managed converged network, tightly integrating data, VoIP, and video services may soon prove to be a huge differentiator between competitors in business. The next chapter will dig into configuration, fault, and performance management.

Download Additional eBooks from Realtime Nexus!