

Realtime
publishers

"Leading the Conversation"

The Definitive Guide[™] To

Converged Network Management



Ken Camp

Chapter 8: Effective Network Configuration, Network Fault, and Network Performance Management.....	162
Fault Management	162
Network and Fault Management—An Integration Strategy.....	165
Caveats of Implementation Strategies for NMSs.....	165
Recognize, Isolate, and Detect Faults	166
The Role of the NMS.....	167
SNMP and Fault Management.....	168
Syslog and Fault Management.....	170
Fault Management and ROI.....	170
Configuration Management	171
Collecting and Storing Configuration Data	172
Configuration and Change Management	172
Performance Management	174
QoS and Bandwidth Monitoring.....	179
Collecting and Analyzing the Data.....	181
Monitoring the Health of the Network.....	181
Performance and Utilization Trends	181
Administration Management for Performance and Planning	182
Gathering Usage Statistics	182
Managing Backups and Synchronization for Performance	182
Summary	183

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 8: Effective Network Configuration, Network Fault, and Network Performance Management

Chapter 7 introduced the FCAPS model and examined service availability and capacity planning management. This chapter will continue that theme of using a methodology for consistent management of network faults or problems, configuration of network devices, and performance.

Network management means different things to different people. For some organizations, it simply means a network consultant is monitoring network activity with some tool. In larger enterprises, network management involves continuous polling of network devices to monitor status, distributed databases containing logs and error reports, and graphical representations of the network topology to present a high-level view of the overall health condition of the network. All network management can be viewed as a service that uses tools, devices, and software applications to assist network managers in monitoring and maintaining the quality of service (QoS) being provided.

Fault Management

As a part of holistic network management, fault management is the term used to describe the set of tools and functions used to detect and isolate and then remediate problems in the service network. These malfunctions may be technical, such as equipment failures, or caused by human error. The central theme is that something failed in the network. Fault management sometimes includes environmental control systems or monitoring.

Faults are detected through monitoring system events. In many organizations, event monitoring occurs in three ways:

- The Network Management System (NMS) in an enterprise command center monitors the status and health of network elements. Often an icon representing an element in the network will simply turn from green to red, indicating a problem. The NMS typically also functions as a fault management system.
- Event correlation and analysis systems are designed to process syslog and event log files from a number of systems. Many of these systems include an engine for detecting anomalies as part of event correlation.
- Human analysis, while often effective, is also inefficient. The sheer volume of log data generated in large networks makes human analysis impractical for many purposes. Yet when network performance issues arise, most groups have people begin to review logs. Humans apply different logic than NMS and event correlation engines apply, and may of those systems spot trends, patterns, or unique items that fall below automated thresholds.

When a failure or fault occurs, elements of the network send information about the problem to the NMS. SNMP is widely used for this purpose. Elements transmit alarm information or indicators. These alarm indicators remain in alert state until the error condition or problem is fixed.

Many fault management systems are configurable to allow prioritization of network faults into different levels of severity. Some alarms may simply represent debugging information, while others might indicate emergency problems. The following list highlights the most common severity levels used in fault monitoring and management systems today

- Warning
- Minor
- Major
- Critical
- Indeterminate
- Cleared



Note the “cleared” severity level. It is considered good practice to have network elements send positive alerting information when an error has been cleared to ensure all monitoring systems note problem resolution.

The fault management console, often referred to as the “dashboard,” allows the operations center staff to monitor events across the network from a large number of systems. A large enterprise network approach might be to overlay the network map on top of a geographic map to give a sense of where in the network events are occurring. A fully robust fault management system may be automated to the point that corrective action is handled automatically. Although many network elements include protocols for automatic failover and resilience, other elements may require creative approaches. For example, detection of a failed server at the NMS might trigger running a script that moves active services to a backup system. Scripting reconfiguration of this type of response helps ensure quick activation and eliminates the potential for keyboarding input errors.

In many cases, the automated system will simply be used to alert network support staff. Today, email messages, pager alerts, and SMS messages to mobile phones are widely supported approaches used in the NMS. It may also be prudent to implement practices that include escalation mechanisms. Notifying a single person is probably inadequate for a very large enterprise network. Sometimes a group of supporting personnel is notified as part of the incident response team.

At a high level, fault management can be performed in active mode or passive mode. SNMP monitoring and data gathering are a passive approach. Data is gathered as part of the normal flow of traffic. If the network element recognizes some error condition exceeding thresholds, an error message is sent to the NMS. This also presumes the network node is able to send the appropriate alarm. If, for example, a power supply were to fail, often no alarm can be sent. The element simply powers down.

Active fault management engages other tools to monitor the health of systems. These tools include PING, Web page retrieval, testing of email responses, port scans, and a number of other methods. If a Web server fails to serve up a Web page, an active monitoring system can send an alert based on that failure. The server hardware might well be fully operational with no detectable malfunction. In this case, perhaps Web services failed, but the server is still operating. No hardware components have failed. Even the OS and Web application may be running properly. Active monitoring can help catch faults that would otherwise pass undetected.

One key to successful fault management is to deploy tools that support processes rather than the reverse. Established processes shouldn't be bent to fit new tools without careful consideration. On the flip side, tools may lead to the development of new network management processes where none previously existed. When deploying network management tools, it's helpful to focus on the fault tolerance and resilience of the service delivery network, with an eye on both redundancy and security.

Network and Fault Management—An Integration Strategy

There are some basic rules of thumb to follow when integrating the NMSs into the converged services network:

- Identify the best sources for data. Not every network element warrants monitoring, but it's vital to monitor key service delivery elements.
- Use the best tools available and the best features of each. A mediocre tool might seem cost effective initially, but if it doesn't provide the necessary capabilities, it won't help maintain an optimal network for delivery of integrated data, voice, and video.
- Wherever possible, use the dashboard approach to deploy a "Manager of Managers." This centralized dashboard will provide a complete holistic view for at-a-glance snapshots of the overall network status and health.
- Follow the tried and true KISS (Keep It Simple Stupid) approach wherever possible. Don't introduce undue complexity.
- Use modular management components that integrate well together. Stay focused on out-of-the-box functionality. An NMS that requires customer script development and programming to implement is likely to prove a problem to sustain. The GUI approach is easier to learn and adopt than a system using command lines. Where practical, specific workgroup GUIs can provide individual teams with the tools they need from a single source (the dashboard or "Manager of Managers"). The network support group, Help desk, voice services staff, and security administrators may all need slightly different GUI views of the same underlying network infrastructure to operate effectively.

Caveats of Implementation Strategies for NMSs

When deploying the NMS, there are some basic caveats to keep in mind. A complex and convoluted system can be less effective and more labor intensive than a manual system of work procedures. Again, keep it simple.

Avoid over-automating. System automation is a wonderful boon to productivity, but the delivery of integrated data, voice, and video services requires human oversight. Automating in the wrong place can lead to a false-positive system event that triggers automated network reconfiguration or failover to backup systems. Look carefully at these mechanisms to ensure that qualified support staff is in control. Remember these are tools for staff, not replacements.

When evaluating NMS solutions, here are some basics to consider:

- Does it have a simple interface? It should be easy to access everything you need. Users shouldn't have to toggle back and forth between screens to perform basic tasks. A Web browser based interface may allow for easy customization for different workgroup views.
- Does it provide the ability to set a baseline? Simply put, the value in an NMS is in the ease of setting baseline thresholds at normal network performance and operation levels. How easy is it to establish the baseline and set thresholds than trigger notifications?
- Does the reporting capability meet your needs? If the NMS can report an event, whether it's a failed element or a spike in traffic, does it also provide enough historical tracking and analysis capability to provide management reports that will be useful to the service delivery manager? An NMS tool that requires knowledge of another programming or report writing language will make it more difficult to extract useful information in a format that helps make informed business decisions.

 Later, this chapter will talk about configuration management, which is tightly coupled with change control processes. Throughout this chapter, the goal is to maintain a broad holistic view, keep perspective, and link all the various components of network management together.

Recognize, Isolate, and Detect Faults

Past studies have given way to anecdotal evidence in many areas of networking. One example is commonly referred to as the “80-20 Rule.” Sometimes it seems that there is an 80-20 Rule to fit every situation. Fault management is no different. The commonly held belief, supported by past studies, is that when faults do occur, 80% of the time is spent looking for the problem, while only 20% of the time is spent fixing the problem. Recognizing that faults exist, detecting them, and isolating them is also referred to as anomaly detection and correction. What we're striving for is near real-time event correlation information to gather a timely root cause analysis of the fault for timely remediation of the problem.

NMS vendors provide this root cause analysis functionality in a number of ways. Some use mathematical/statistical modeling techniques. Others base solutions on complex academic systems models. Many take a much simpler approach using rudimentary event filtering schemes. This last category is often the approach implemented by fault monitoring systems that organizations develop in-house.

The Role of the NMS

Regardless of the methodology used internally in the NMS, event filters sort through the flood of SNMP traps that are constantly being processed. In its most basic form, which Figure 8.1 shows, the NMS provides two key roles—correlation and an event watcher. Most organizations configure the NMS to ignore trivial events and closely monitor areas of concern. As events are processed, they're typically handled in one of the following ways:

- Traps reporting many failures are analyzed, leading to a verification, event correlation, and a notification process. These are typically warnings and minor alarms.
- Traps reported “clear” or restored may also trigger notifications to cancel other work tasks.
- Some traps, especially those deemed as major or critical, may trigger immediate notifications.
- Information and debugging traps might simply be logged for future reference

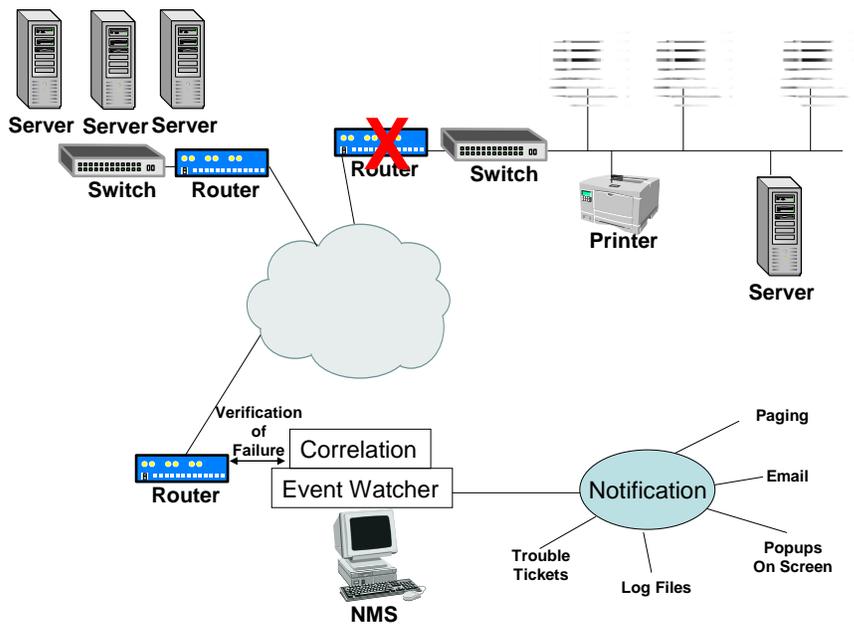


Figure 8.1: A basic NMS.

SNMP and Fault Management

Chapter 6 reviewed SNMP, which is a crucial protocol in holistic fault management and network health monitoring. SNMP is the most widely used protocol for managing devices in the network. It's popular because it's flexible and easy for vendors to implement. SNMP management includes three components: managed devices, agents, and the NMS.

A managed device is any hardware element in the network that implements SNMP and is capable of reporting information. This includes routers, switches, servers, workstations, printers, and other devices. In the converged services network, the media and signaling gateways, voicemail systems, and other call processing elements will all probably provide SNMP monitoring and management capabilities. They may also include vendor proprietary mechanisms.

An SNMP agent is simply the software that provides the SNMP information. The agent might be a *daemon* process running within the OS kernel or some additional software that is installed at the time the network device is set up. The agent software collects information and passes it to the NMS.

The NMS is the centralized monitoring overseer. The NMS sends requests to monitored devices in the network, and the agent software running on the device sends a reply. The NMS sends five types of messages:

- GetRequest is used to retrieve a specific value from a network device.
- SetRequest is used to set a defined value in a network device.
- GetNextRequest is used by the NMS when building a table of responses. It's used to collect multiple inputs.
- GetResponse is used to return error codes and other responses to requests from the NMS.
- Trap is an unsolicited message. It's sent from the agent to the manager. This is the error reporting mechanism that is used to provide immediate information to the NMS about the status of a network device.

 For technical details on SNMP see the following resources.

The SNMP Version 1 RFCs are:

- * RFC 1155. Structure and Identification of Management Information for TCP/IP-based internets
- * RFC 1157. Simple Network Management Protocol
- * RFC 1212. Concise MIB Definitions
- * RFC 1213. Management Information Base for Network Management of TCP/IP-based internets

The SNMP Version 2 RFCs are:

- * RFC 1901. Introduction to Community-based SNMPv2
- * RFC 2578. Structure of Management Information Version 2 (SMIv2)
- * RFC 2579. Textual Conventions for SMIv2
- * RFC 2580. Conformance Statements for SMIv2

The SNMP Version 3 RFCs are:

- * RFC 2576 (PROPOSED STANDARD). Coexistence between SNMP Version 1, Version 2, and Version 3 (March 2000)
- * RFC 3410 (Informational). Introduction and Applicability Statements for Internet Standard Management Framework (December 2002)
- * RFC 3411. An Architecture for Describing SNMP Management Frameworks (December 2002)
- * RFC 3412. Message Processing and Dispatching (December 2002)
- * RFC 3413. SNMP Applications (December 2002)
- * RFC 3414. User-based Security Model (December 2002)
- * RFC 3415. View-based Access Control Model (December 2002)
- * RFC 3416. Version 2 of SNMP Protocol Operations (December 2002)
- * RFC 3417. Transport Mappings (December 2002)
- * RFC 3418. Management Information Base (MIB) for the Simple Network Management Protocol (SNMP) (December 2002)
- * RFC 3584. Coexistence between Version 1, Version 2, and Version 3 of the Internet-standard Network Management Framework
- * RFC 3826. The Advanced Encryption Standard (AES) Cipher Algorithm in the SNMP User-based Security Model

Syslog and Fault Management

As described in Chapter 6, syslog is another important tool in fault management. SNMP is widely used for status and health monitoring. It provides real-time information about elements in the network.

Syslog provides more comprehensive log data, including error messages with extensive data about what is occurring internally within the network device. As syslog data comes from an array of devices throughout the network, it's used in event correlation to understand the impact a failure in one part of the network may have in other areas.

Because syslog data is in plaintext, it's easily readable, even when not easily understood. This makes it easy to manipulate for analysis. Syslog data is crucial for verification and validation of details associated with many network faults.

 Chapter 9 will explore the value of syslog information in assessing security issues.

Fault Management and ROI

There are tangible and measurable return on the investment (ROI) values in fault management methods. The level of service delivered to users is higher when fault management and monitoring are applied. NMS tools provide immediate impact assessment capabilities for response to faults and problems arising in daily operations. This leads to better decision making, both in troubleshooting and network planning. This improved operational stability will result in fewer outages and errors and improve the QoS delivery overall. Consistency in service delivery engenders user confidence in the integrated services. Satisfied users will leverage the converged services into other enterprise business process flows. Knowing and understanding the details of problems in the service network provides not just historical information but also crucial business intelligence that aid in future capacity planning and new services for consideration in the future.

Configuration Management

The primary objective of configuration management is to monitor network and system configuration information to provide an audit trail, or tracking mechanism, of changes made to the network. Every device in the network has some associated configuration information. Given the variety of network elements in the converged services network, it's critical that configuration changes be coupled tightly with some formalized change control process.

As with every other aspect of managing the FCAPS model, the NMS is the brains of the overall process. NMS solutions might be single-vendor, single-system solutions or they might be distributed systems that work together and roll information to the previously mentioned "Manager of Managers." NMS is a system that may have several components.

Some key features to consider when investigating configuration management include:

- **Auto-discovery**—Is the system capable of automatically probing your network and learning what the components of the infrastructure? Can it determine topology and layout of the existing network, and detect new elements as they're added? Can it be configured to automatically retrieve configuration information and store it in the configuration management database?
- **Ability to import configurations**—Will the system support all the various configurations of different vendors' solutions into a single database schema? Will different configuration databases be required for servers and workstations than for switches and routers? Are all the elements of the converged services network supported (call managers, media/trunking gateways, signaling gateways, voicemail systems, and traditional telephony elements such as PBXs)?
- **Configuration analysis**—This is a new development in NMS capabilities that has arisen in the past few years. Is the NMS able to compare existing configurations with other similar devices in the network? Does it include a standardized comparison of known best-practice configuration issues? Routers may have common configuration errors that don't impact prior operations but might impact converged services. Servers and workstations may have unnecessary services running that should be disabled. OS versions and patch levels need to be monitored, and in most networks, maintained with consistency across similar network devices. Will the NMS provide this comparison capability?
- **Policy-based configuration**—Does the NMS support a single policy-based solution and provide for easy deployment across all devices in the network? Policy-based configuration management allows precious staff time and resources to be used more effectively. The ability to write a policy template that can then be universally "pushed" ensures consistency of operations and unifies compliance with enterprise standards.

Collecting and Storing Configuration Data

In unified communications, network configuration management provides control of changes in the network. The most common changes noted are to hardware, firmware, and software, but documentation changes also apply. And these changes continue throughout the lifespan of the network. Change is constant, driven by upgrades, patches, and equipment replacement. A comprehensive configuration management approach will collect all the changes made to a network element throughout its entire life cycle.

Once a device is deployed in the network, change control processes guide the evaluation and approval of implementing configuration changes. This function is vital for two reasons:

- The interrelationships between network elements become more complex as the network incorporates new services. In the past, a simple change to a router might have only impacted traditional data flows. In the converged network, a simple router change can easily alter the QoS for a traffic flow. This simple change might disrupt or degrade voice services unexpectedly.
- Fallback to previous configurations may be required quickly. One of the great benefits of comprehensive configuration management is the ability to undo changes and revert to a prior, known operational state.

Configuration and Change Management

Configuration and change management thought processes in the industry today are loosely based on guidelines for software configuration management principles described by Roger Pressman in his book *Software Engineering: A Practitioner's Approach* (McGraw Hill, 2004, ISBN: 007301933X). SCM is a widely accepted methodology for controlling and managing change in the software development environment, but the core principles apply equally to networking. The central focus is to identify what changed, who changed it, and how the change can either be reproduced or undone.

Configuration Management Resources

The Institute of Configuration Management at <http://www.icmhq.com/index.html>

Configuration Management Training Foundation at <http://www.cmtf.com/>

Configuration Management Body of Knowledge at <http://www.cmbok.com/>

The Configuration Management Wiki-Web at <http://www.cmcrossroads.com/cgi-bin/cmwiki/view/CM/>

Configuration Management Principles and Practice by Anne Mette Jonassen Hass (ISBN: 0-321-11766-2)

Configuration management provides for the identification of changes, a controlled storage environment for archival data, change control, and status reporting of every change activity throughout the life cycle. Figure 8.2 shows a simplified view of configuration management activity areas and offers a peek into process flow. The configuration data itself is shown as metadata (data about the configuration information). In configuration management, the metadata for a configuration item may include:

- The name of the change
- The name of the person initiating the change
- The name of the approver of the change
- Text description of the change
- Date change was placed into production
- References to other configuration items

In most NMS solutions, the metadata is often stored on the same system as the configurations themselves.

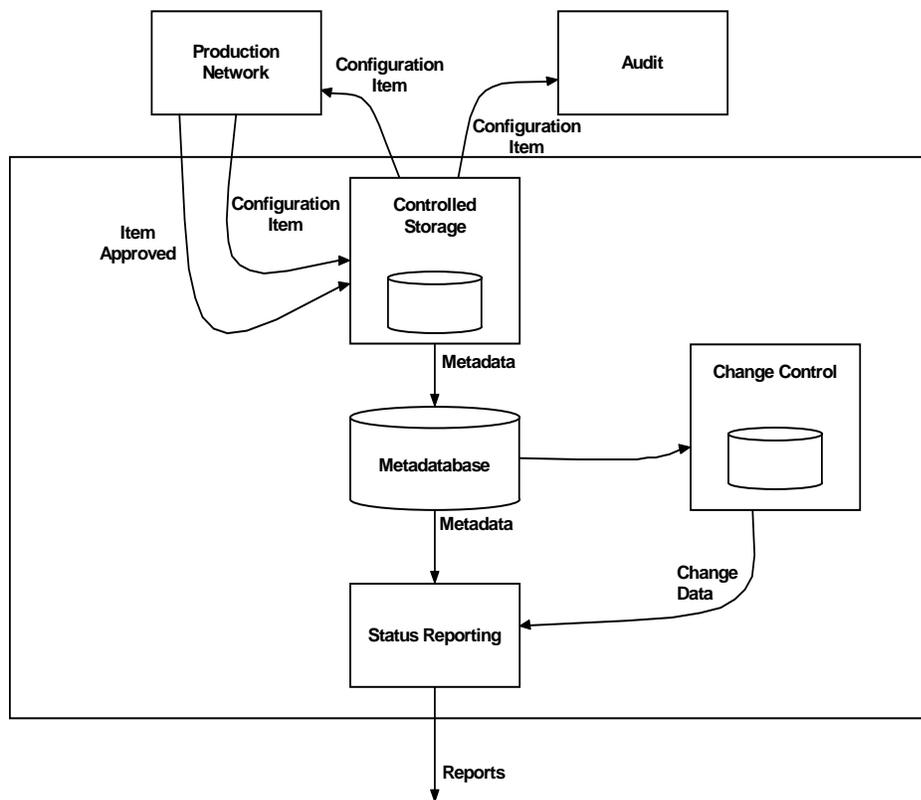


Figure 8.2: Configuration Management Activities.

Configuration items that are different versions of the same original item are obviously strongly related, but each one is an individual item, which will be identified and may be extracted and used independently. This is one of the main points of configuration management: to be able to revert to an earlier version of an item class.

At every step of the FCAPS methodology there is a documented, methodical process being reinforced as a commonly accepted best practice. In Chapter 7, Figure 7.7 shows an evolutionary path of an enterprise network moving from anarchy through reactive and proactive modes as part of the maturation to a service-oriented business enabler. The maturation cycle of a network often mirrors the evolution of the business itself. Configuration management is a vital part of collecting and managing business intelligence about the services supporting the core business.

Performance Management

One key to the emerging integrated services network is the constant appearance of new technologies. Today, 10 years into deployment, VoIP isn't really new or emerging. Within the broad telecommunications sector, many service providers are now stepping back and thinking more purely in terms of voice services, with VoIP being one voice delivery mechanism.

For enterprise networks, the impetus is much greater to integrate technologies with business processes. You'll recall that Chapter 1 looks at convergence from several views:

- Infrastructure convergence of wiring and circuits
- Service convergence using IP as the delivery mechanism
- Device convergence of the physical tools used in everyday business
- Application convergence of enterprise business applications such as sales force automation (SFA), supply chain management, and customer relationship management (CRM)

Although commercial providers of voice services might be starting to take a narrower view of voice as simply a service, that is a luxury the enterprise business cannot afford. The value in convergence is found in the total integration. Complete integration can reduce cost and increase productivity. Once integration has become inculcated into the corporate culture, it can lead to creative new solutions to old business problems and even spawn ideas for new business tactics.

Chapter 5 looked at the Network Performance Envelope methodology for assessing requirements. One tangible benefit to this model is that it supports many phases of the life cycle of the enterprise converged services network. Figure 8.3 touches on some of the most basic principles of this model.

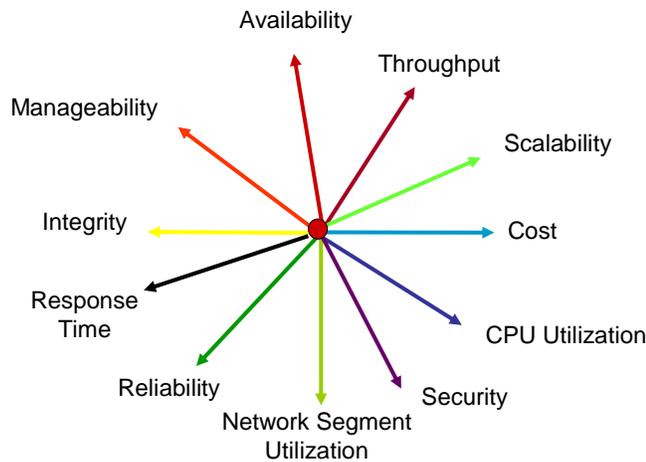


Figure 8.3: Revisiting the Network Performance Envelope.

There are many factors that impact performance of the network. Prior to implementation, you assess network readiness—you determine whether the network can support converged data, voice, and video. As you implement those services, you reassess to ensure that your assumptions have all been proven correct, that any requisite upgrade activity was successful, and that the network does indeed deliver the integrated services as planned.

In the operational phase, you continually monitor performance to ensure those standards are being upheld. Operational success is driven by information. The more you know about the network, the better your service delivery consistency. Although network managers all strive to deliver *good* service, sometimes consistency of service gets overlooked. Vigilant performance monitoring helps ensure a consistent, reliable network.

Performance management is a series of checks and balances. We often visualize a counter-balance scale in weighing performance against other factors. That simplistic view doesn't serve the needs of managing a converged service network. Some managers view this multi-service network in terms of the trade-off triangle (see Figure 8.4).

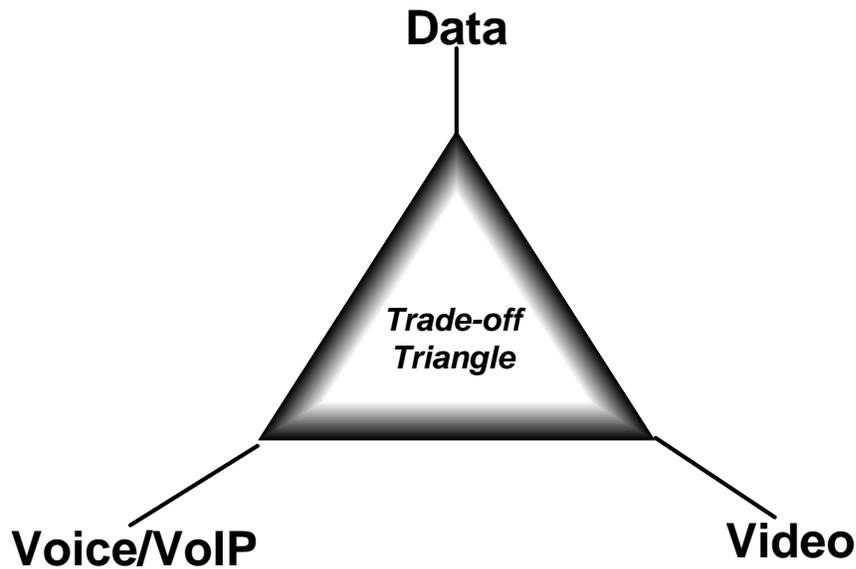


Figure 8.4: *The misleading trade-off triangle.*

The danger in taking this view of balancing performance and service is that it pits network services against one another for resources. This approach may be simple, but the danger of giving preference to one of the new emerging services, VoIP or video, is that existing data services—Web services, financial applications, customer service tools—may suffer from an exhaustion of resources. Business-critical applications may starve for resources while the new technologies take center stage. The holistic truth is more akin to balancing a high-performance automobile tire.



A binary decision of either this or that will lead decisions about holistic network management to ineffective conclusions

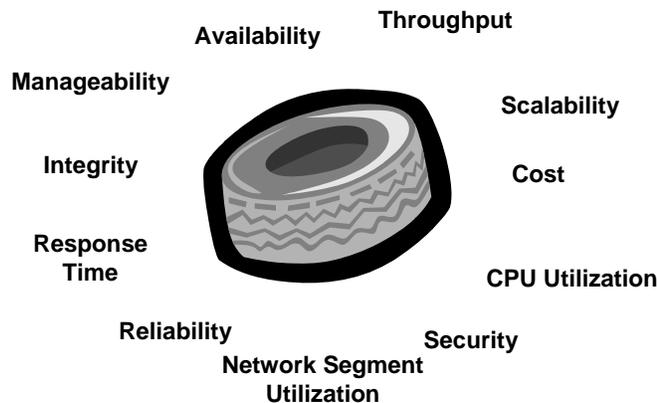


Figure 8.5: Managing holistic balance.

The illustration in Figure 8.5 has simply taken the network performance parameters used in the performance envelope methodology and shown them as balance points around a tire.

Performance monitoring and management is not a binary decision in any way. Trying to balance security against manageability, error rate against response time, or throughput against reliability are just a few examples of how binary decisions fail.

Because much of the focus at this stage of convergence is on VoIP services, look first at the VoIP service elements that must be monitored. Remember that years of telephone usage have socialized us to expect a base performance and quality profile akin to the traditional telephone network.

When VoIP devices register with the network, whether they are phones, media/trunking gateways, signaling servers, or other devices, registration problems can adversely impact the delivery of service. If a VoIP phone on a desktop can't pull down a profile and phone number for any reason, problems follow. You'll want to establish a monitoring and alerting threshold in the event that something as simple as registration attempts or failures exceeds predefined levels. If there is growth in the network, or the number of registered telephones changes dramatically, it could be a signal that there is a problem with the VoIP services themselves. Gateway registration monitoring will help identify new or missing trunk capacity to other networks.

Call monitoring is often an inflammatory topic because of privacy concerns. At this level of providing service in the network, call monitoring doesn't mean eavesdropping on individual calls. It's truly call traffic monitoring. This function involves monitoring incoming and outgoing call volumes to identify failures. If the VoIP system supports fax calling, attempted fax calls should be monitored as well. In call monitoring, look at four distinct areas:

- **Calls in progress**—The instant a VoIP phone or softphone goes “off hook,” the call is deemed in progress until the caller or called party hangs up. Calls in progress include those in the process of dialing, off-hook getting busy signal, and so on. If every call attempted completes successfully, the number of calls in progress will be the same as the next time, active calls. As part of capacity planning, decisions have to be made at implementation time as to the number of active calls the system can support. Performance monitoring assists with capacity planning. If the percentage of available capacity runs normally at 70% utilization, then rises to 95%, performance monitors can alert capacity planners to a change in the network; a change that must be addressed before calls are blocked due to demand exceeding the capacity.
- **Active calls**—Once a call has successfully connected, it's deemed an active call. It has completed the setup signaling and now has a voice media path connected through the network. The same capacity considerations apply as for calls in progress.
- **Attempted calls**—In general, designers try, within available resources, to provide non-blocking systems. That is, they try to guarantee that all calls attempted will be completed successfully, but such isn't the case in the real world. Calls encounter busy conditions and go unanswered or are abandoned. Monitoring calls attempted over time yields data used to identify peak calling traffic periods and what is referred to as the busy hour call attempt (BHCA) value.
- **Completed calls**—Any phone call that completes and terminates without an abnormal termination code is counted as a completed call. Monitoring completed calls over time is also helpful in determining peak traffic periods and the BHCA value.

For most businesses, using VoIP strictly as an internal calling system isn't practical. Corporate VoIP services have to interconnect with the Public Switched Telephone Network (PSTN) through either SIP trunks or standard voice trunks. This is done through some form of gateway. A gateway connecting to the PSTN via traditional T1 trunks needs to be monitored on the PSTN side of the VoIP service network not just on the IP network. Monitoring active channels on these trunks over time can help identify calling patterns and busy hour peak call volumes. Baseline data can also be used to identify underutilization of circuits, leading to downsized capacity and reduced operating costs. Data trending helps in capacity planning and the growth and maturation of the converged services.

Another benefit to implementing VoIP services is the conference-bridging capabilities. Conferencing adds yet another monitoring aspect. If your VoIP implementation supports conferencing, the maximum number of audio streams that can be supported for conferencing has to be configured. Monitoring will help ensure that the number of available audio streams meets the service levels needed for day-to-day business requirements.

IP phones, whether they're physical hardware phones or softphones running on workstations, require continual monitoring for service assurance. IP phones should be monitored for registration status, dial tone validity, jitter, latency, and packet loss. These key QoS metrics directly impact both call quality and the user experience.

QoS and Bandwidth Monitoring

Earlier chapters talked about codecs and QoS. Bandwidth requirements for VoIP traffic are driven in large part by codec selection. The PCM codec (G.711) requires about 64Kbps to support a bi-directional phone call. The G.723 and G.729 require significantly less bandwidth due to compression, but network congestion may have a greater impact on call quality.

Whenever new applications are introduced into the business network, there is a risk of oversubscribing individual links. Oversubscription leads to congestion that can, in turn, degrade call quality. Packet loss and increased latency are common side effects of congestion. In the worst case, they can disrupt call setup and voice media transmission to the point that VoIP services become unusable.

To guarantee that VoIP users receive an acceptable level of voice call quality, VoIP traffic generally needs to be given priority over other types of traffic on the network, although video traffic may warrant even higher priority than voice. Data traffic is bursty in nature and often does not involve real-time communications between people. The primary objective of QoS techniques is to provide a prioritization mechanism that can ensure that every type of packet on the network is handled appropriately for the content inside. Because of its real-time nature, VoIP traffic typically receives the preferential treatment to reduce or eliminate delay. Commonly monitored voice performance metrics include:

- **Delay**—Latency or delay is an estimate of the packet delivery time across the network. It's expressed in milliseconds and measured as the average value of the difference between a sender's and receiver's timestamps on messages. It's measured when the messages are received. Remember that delay is cumulative. End-to-end delay is a key factor in determining the overall VoIP call quality.
- **Jitter**—Variation in delay is called jitter. We've mentioned consistency in performance, and jitter is one area that is crucial in delivery of real-time services such as voice and video. It indicates a variance in the arrival rate of packets at the destination. Jitter is a predictability factor that is often used in discussion of the overall reliability of a service network. Jitter problems are well known to adversely impact call quality. Networks can compensate for jitter by implementing jitter buffers to normalize the timing of the traffic flow. Jitter buffer loss occurs when jitter exceeds that which the jitter buffer can hold. Jitter and jitter buffer loss affect call clarity, which affects the overall call quality and user experience.

- Packet loss—Loss simply indicates packets lost or discarded in transmission. In VoIP services, this could mean the loss of an entire syllable or word during the course of a conversation; more importantly, it might mean the loss of a dialed digit, preventing a call from ever completing successfully. Obviously, data loss can severely impair call quality. Monitoring systems measure the number of packets that were expected against the number actually received.
- Mean Opinion Score (MOS)—MOS is a subjective quality measure that was discussed in earlier chapters. MOS testing is non-intrusive and provides a way to monitor and measure call quality in ongoing network operations. Historically, MOS was derived from panels of judges listening and scoring call quality. Although it's useful for human users to interpret quality, it's a useless metric in delivering quality assurances under any form of Service Level Agreement (SLA). It doesn't fit well in operational monitoring systems; but there are alternatives that are gaining popularity, especially with large enterprises and service providers.

The E-Model is a tool that can predict the average voice quality of a call using a mathematical model. E-Model accounts for the estimated impact of delay, jitter, loss, and codec performance. The output result of an E-Model calculation is called the *R Factor* or *R Value*. These values estimate voice quality on a scale from 0 (the lowest quality) to 100 (the highest quality). Like an MOS score of 5, an R Value of 100 is, in theory, unattainable, but it's the standard goal network engineers shoot for when designing voice service networks.

Because E-Model scores are based on measurable parameters, monitoring tools are becoming more common that can be incorporated into the enterprise NMS strategy for performance monitoring. E-Model monitoring tools evaluate the Real-Time Protocol (RTP) streams based on information found in the traffic flows (source address, destination address, TCP/UDP port numbers, and packet sequence numbers) to create what is called a "jitter profile." E-Model then creates a score that, in testing, correlates to traditional MOS with 80 to 90 percent accuracy.

There is another non-intrusive performance measurement technique that is emerging called the ITU-T Calculated Planning Impairment Factor (ICPIF) score. It's based on the ITU-T G.113 standard. Driven by increasing sophistication required to accurately assess voice quality, some vendors are beginning to adopt this technique. With increased attention to VoIP quality in standards bodies such as the IEEE and ITU-T, this method, which has been around for nearly 10 years, is gaining traction across the industry.

ICPIF takes some factors into account that the E-Model does not:

- Signal attenuation distortion
- Circuit noise
- Codec encoding distortion
- Delay distortion
- One-way transmission time
- Echo

Although ICPIF may be slow moving into enterprise business network monitoring, it will certainly play a role among service providers in delivering commercial voice services using VoIP.

Collecting and Analyzing the Data

Throughout, this guide has repeated a common theme: Information gathering is absolutely necessary to successfully manage a converged service network. The more extensive and comprehensive the data collection and analysis tools, the better armed network managers are to make decisions about present daily operations and future capacity and growth plans.

Monitoring the Health of the Network

Ongoing health monitoring provides the service delivery organization information with which to make informed decisions. Perhaps more importantly, it provides a reportable mechanism to demonstrate service delivery assurances to end users and customers.

Performance and Utilization Trends

Close monitoring of network services brings the tools to hand to proactively measure performance and utilization for multiple reasons. The enterprise network—particularly the integrated services network providing delivery of a mix of data, voice, and video services—is a finely tuned machine. Just as a high-performance vehicle will get better mileage, achieve faster acceleration, and handle better on the race course, a finely tuned network will handle peaks in traffic and unplanned events more smoothly than a network just left to run.

The dashboard of the NMS provides a window into the performance levels and utilization of the network at any point in time. Changes in operating conditions can be detected quickly and proactively, ensuring that business-critical network services are always available.

Administration Management for Performance and Planning

In addition to daily operations, monitoring all these factors provides trending information on utilization, traffic patterns, and capacity. This enables network planners to accurately anticipate future network needs. This added knowledge factors into business decision in other ways.

Knowing that capacity planning needs are being closely managed helps ensure that a company doesn't overspend in future capacity that isn't required to support the business. Information acts as a quality assurance mechanism in future network planning.

Gathering Usage Statistics

Beyond the call quality and user experience measures and metrics, the infrastructure itself needs to be monitored closely to ensure the critical service delivery elements are operating at appropriate levels and can handle the load of traffic being passed. Minutes of use in the traditional telephony environment were used as an indicator that enough T1 trunks and circuits were available. In the converged network, minutes of use correlates to bandwidth and network resource consumption. A trend indicating growing minutes of usage means more network resources are being used. It can provide an early indicator that SIP or TDM gateways, signaling servers, voicemail systems, and other elements of VoIP service delivery need to be evaluated to ensure they aren't being stretched beyond tolerable capacity levels.

Disk space and CPU utilization are very simple indicators to watch and good barometers of changes in the environment. If any vital network element spikes in CPU utilization, or suddenly runs out of disk space, monitoring helps ensure an early warning system can engage immediate response before service delivery is impacted.

Managing Backups and Synchronization for Performance

In enterprise business, backup systems are widely deployed to ensure all mission-critical data is protected. When delivering converged services, some new considerations may come into play. It's imperative that you remember to back up not just the data systems but also the management systems. The NMS in an integrated network is a mission-critical piece of service delivery inside the enterprise. Backup and restoration plans need to be developed and rigidly followed. After investing all the time and effort required to baseline the network, collect configuration information, and build a historical library of past performance metrics and trends, it would be disastrous to lose the information simply because the NMS wasn't part of the information backup strategy.

For large enterprises and geographically dispersed organizations, continuity of operations may involve a different approach. Many companies have a business continuity or disaster recovery center located at a remote location. As part of the NMS strategy, it's wise to consider a mirrored NMS at the remote facility to ensure uninterrupted service in the event of a disaster.

Summary

Networks are increasing in both breadth and depth. The increase in breadth relates to an increase in size. More users connect to enterprise networks. The number of connected devices is also growing. There is a parallel increase in depth driven by the set of services delivered to any endpoint on the network. In earlier networks, the service delivery footprint was small, relegated to FTP, email, and Web browsing. Converged networks are rapidly changing that. The integrated network is now bringing data, voice, and video together. From an infrastructure perspective, that drives the need for more vigilant monitoring of faults, capacity, assets, performance, and security.

 Chapter 9 will address security in the converged services network in more detail.

The converged network is also rapidly growing in other ways. Voice and video are viewed as network services. Application services are also quickly integrating, both into the network and with voice and video. There are multiple concurrent convergences underway, making it vital that you monitor closely to protect your business services. Providing users with a consistent QoS and reliable network experience in increasingly complex networks requires the use of increasingly sophisticated tools.

Legacy NMSs don't do a good job of managing emerging, converged networks. The old tools focused on a GUI interface with a network status indicator. In today's multi-service network, that may not be enough. New tools are based on algorithms for network discovery, real-time monitoring, visualization tools for modeling, simulation, anomaly detection, and event correlation.

This guide has touched on several views of life cycle management. Figure 8.6 presents another—a simple network performance management life cycle.

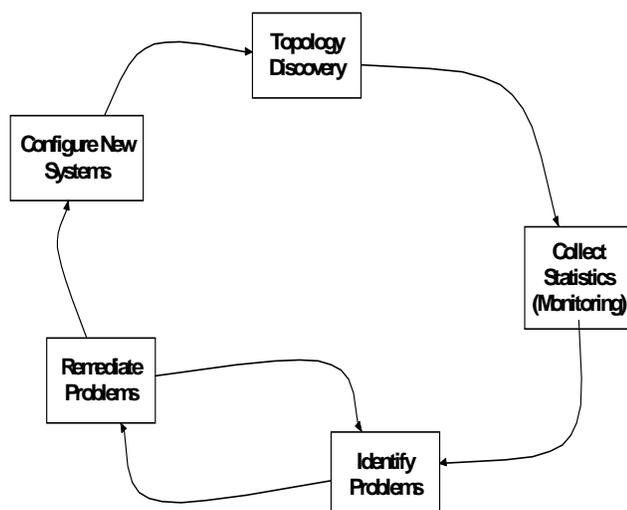


Figure 8.6: Network performance management life cycle.

Many legacy NMS approaches only deal with discovering network topology and collecting statistics. When planning for the ongoing life cycle management of an integrated data, voice, and video network, it's important to look at NMSs that can incorporate problem identification and remediation along with the configuration of new systems.

 As a note on topology, remember that with current networking equipment, topology is more than just ports and cables. The logical topology of VLAN configuration, MPLS domains, and IP subnetting is an important facet of topology discovery in depth.

As networks evolve, QoS guarantees for bandwidth and delay characteristics for audio are a starting point, but they aren't enough. The technologies are advancing rapidly and this is a good time to consider broader performance attributes such as variations in connectivity, robustness of redundancy and failover, and the overall performance of the traditional best-efforts approach used by IP. For future networks, these tools will need to provide for:

- Performance monitoring and measurement
- Forensic analysis (beyond packet sniffing)
- Capacity planning (to project impacts of new services)
- Load generation (for scenario testing)

The only way to can achieve an ideal NMS is to proactively incorporate components of performance monitoring into the system. Metrics for availability and responsiveness need to be included in any notification system. Baselining support is intrinsic to success, and routinely scheduled baselines should be run across the corporate network.

Without an NMS that is oriented to the complete array of services provided, including data, voice, and video, companies will fly blind as to service assurance. VoIP services that lack comprehensive management and monitoring procedures are likely to deliver poor quality services. Continual monitoring of documented service-level metrics (appropriate QoS metrics) is the only way to truly ensure and demonstrate enterprise-class service delivery. These metrics should include, at minimum, some combination of jitter, latency, call completion and quality, and voice quality (MOS, and so on).

Chapter 9 will wrap up the dive into the FCAPS model and will highlight security management issues facing enterprise deployments today. In addition, the chapter will identify common industry best practices for creating an effective VoIP security plan that balances securing the network against the VoIP requirements for availability, reliability, and performance.

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>.