

Realtime
publishers

"Leading the Conversation"

The Shortcut Guide[™] To



**Network
Management
for the Mid-Market**

sponsored by



Greg Shields

Chapter 2: Performance Management	23
Key Steps in Managing Performance	23
Performance Baselining.....	24
Monitoring Deviations	26
Performance Reporting.....	27
Performance Correction	28
Key Measurements in Performance Management	29
Bandwidth Utilization.....	29
Network Latency.....	30
Interface Errors and Discards.....	31
Network Hardware Resource Utilization.....	31
Buffer Usage	32
The Business Metrics of Performance Management	33
Availability and SLAs.....	33
Bandwidth Monitoring.....	34
Link Costs	35
Traffic Management and Prioritization.....	36
Additional Tools for Managing Performance	37
Traffic-Generation Tools	37
Traffic-Analysis Tools	38
Wireless Performance Tools.....	40
Performance Affects Business	40

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 2: Performance Management

The real productivity killer in most networks is a performance level that doesn't meet the needs of users. When network performance is consistently below acceptable levels, business cannot operate at full efficiency, workers can't accomplish tasks on time, and the regular movement of business suffers. To exacerbate this situation, smoking out performance issues on a network is virtually impossible without the proper toolset. If you've ever gotten the dreaded "the network is slow" phone call, you know how difficult it can be to track down the problem. This chapter will discuss the tools that can prepare you for when that call comes, enabling you to respond with "I'm on it. I know what the problem is."

The previous chapter outlined the FCAPS model of network management and how that model will be used to guide the conversation on network management fundamentals. It then zeroed in on the F in FCAPS to talk about fault management. That discussion broke down the steps in fault management and talked about the best ways to implement formal and informal tactics in detecting and correcting faults. It also illustrated how implementing an effective network management system (NMS) that provides for monitoring and alerting is the first step in moving from a reactive administration model to one of proactive administration. Chapter 1 dove into four key technologies—SNMP, SNMP traps, MIBs, and Syslog—and how these four technologies are critical for the operation of a successful NMS.

This chapter will build on this foundation discuss how these technologies can be used for the P in FCAPS: performance management. Starting with an analysis of the four key steps in managing performance, this chapter will enlighten you about the items to document, the metrics to monitor, and the actions to take to ensure your network is operating at peak efficiency.

Key Steps in Managing Performance

Although FCAPS identifies performance management at the same level as fault management, one could argue performance management is a subset of fault management. They both involve the identification and elimination of issues on the network that reduce worker productivity. Identification of performance bottlenecks in your network is done with nearly the same troubleshooting and scientific method as the process for identifying faults.

However, where performance management differs is in how it affects worker productivity. Fault management is easy for a business to incorporate because of the natural on/off nature to faults. When a network device incurs a fault, that network device typically "goes down." Non-functioning network devices have a clear need for fixing because they prevent work from occurring. Within performance management, the clear line between up and down grays somewhat because no device is technically down. Performance issues not attended to can linger for an extended period of time, causing a long-term reduction in network and worker productivity.

It is this graying of the problem that makes performance management so difficult to track from a metrics point of view. If the network pipe between your main office and branch office has a 20ms latency, how does that affect worker productivity? What about if that latency grows to 200ms?

Let's take a look at a variation of the key action steps analyzed in Chapter 1. Figure 2.1 describes the four phases of performance management and the actions associated with those phases.

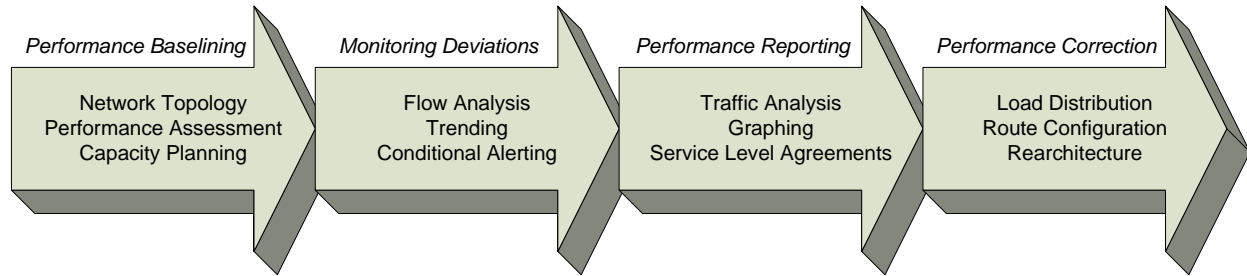


Figure 2.1: The four key steps in managing performance and their associated action steps.

This graphic shows that performance management is really about monitoring for deviations from nominal performance. In the first phase, a baseline of performance is captured using various baselining tools. This baseline is used in monitoring for deviation, and further tools and techniques such as trending and conditional alerting are leveraged to watch for those deviations. Performance reporting is a further component of the deviation monitoring that can provide a graphical representation of performance to the administrator and to business management. All these feed into the correction of the performance issue. That correction may involve a distribution of the network traffic, a reconfiguration of the route, or a rearchitecture of the underlying network.

Performance Baselining

The first step in any performance management activity is to truly understand your network. If you cannot understand your network in the “good times,” you have no comparison for identifying problems during the “bad times.” Plus, knowing what you consider good performance gives you the quantitative metrics to justify or invalidate users’ complaints of “the network is slow.”

There are three major components of performance baselining: Documenting your network topology and the components that make up that topology, completing a performance assessment of the critical applications you want to bring under management, and the use of both of these tools for understanding and planning for capacity needs.

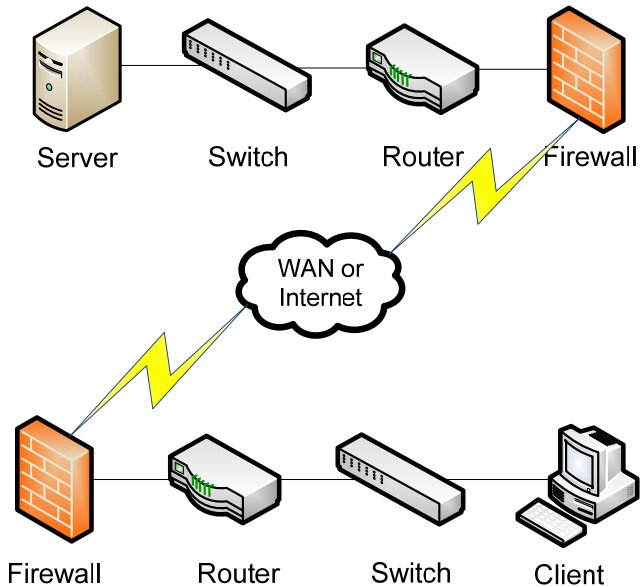


Figure 2.2: Connecting a server to a workstation involves numerous hops from server to switch to router to firewall, through a public network connection, and back through the same devices.

Network performance involves a thread of connections from device to device before the information gets from server to workstation. At each point along that thread, the configuration of the network device and its connection to other network devices will have an impact on the total time it takes data to traverse the network. If even a single connection is misconfigured at a lower connection speed or duplex, it can significantly reduce the total delay.

Effective NMS tools provide the capability for autodiscovery of network devices and their related interconnections. This process of network discovery significantly reduces the time for drawing your network topology. Additionally, a good NMS tool will provide the configurations that explain the lines between the devices to make it easy to spot misconfigurations.

Once the topology of a network is known, the process of completing the performance assessment grows simple. To understand the “good times” of your network, it is necessary to understand what is considered appropriate timing for data to traverse the network. Timing application initiation, data retrieval, and data writing across the network will give you an understanding of the nominal performance of the network. It is within this activity that the concepts of bandwidth utilization and network latency grow important. These terms will be discussed in a minute. But for now, recognize that a time-and-motion study of the network is part of your performance assessment action.

These time-and-motion studies can also be referred to as Network Readiness Assessments (NRAs). Specific to an NRA, you are documenting the attributes of your network to verify its capability to support its hosted applications. In an NRA, you will use the NMS tools described earlier to interrogate the devices on your network to ascertain TCP/IP characteristics such as packet jitter, loss, and delay. These characteristics can affect application response timing across the network both within the LAN and extended throughout your company WAN.

These tasks lead directly into capacity planning. Once you have an understanding of the performance baseline of your network, you can begin to plan for current and future capacity. Most networks are initially architected with consideration for supported applications. But the growth of business invariably adds application support requirements to the network over time. This upward trend in support requirements will add a burden to the network over time, and only through effective capacity management and planning will correct purchasing decisions be made to scale the network. Later, this chapter will discuss some useful metrics that your NMS can monitor and archive that will give you the objective ammunition you will need to justify future network enhancement requests.

Monitoring Deviations

Once you have an understanding of the nominal performance of your network, you can then begin configuring your NMS tool to watch for changes in that performance. This performance monitoring is often done through the use of network probes or network probing. With network probes, hardware devices are installed in-line with devices. These agents or devices report back to the NMS on the traffic characteristics going through the probe. Conversely, with network probing, a centralized device—typically the NMS itself—interrogates each device on a regular interval to pull counter information off that device.

Either technique for gathering data can provide the necessary information you need to monitor for performance deviations in your network. There are some factors in using probes that you will need to consider—such as the need to actually install and administer the probe as well as ensuring that the presence of the probe itself doesn't interrupt the flow of traffic around the network. This additional administrative overhead may drive you towards an agent-less solution.

In either configuration, an effective NMS will provide the capability of regularly gathering data characteristics and alerting when those characteristics exceed administrator-determined thresholds. Similar to how faults are detected, SNMP and SNMP traps are typically used by the NMS to handle configuration and notification associated with those characteristics. Some examples of interesting characteristics are:

- Input/output bits/second
- Current/average response time
- Peak traffic load
- Interface errors/discards
- Percent packet loss

What differentiates a good NMS from an inferior one is the capability to store performance characteristics in a searchable database. This functionality allows the administrator to effectively “go backwards in time” to compare performance characteristics from today's network with those of yesterday or last year. Because network performance issues may not necessarily involve noticeable spikes in traffic, this provides the administrator the ability to do long-term trending analysis on performance. If the use of your network is slowly increasing over time, you can monitor that use over many months to see where your capacity planning needs lie.

Performance Reporting

Obviously, without any reporting capabilities, all this monitoring and database storage of performance results isn't useful. So, an NMS must have a robust reporting engine that can align statistics with the network map and across multiple devices.

When looking for deviations in performance across the network, it is often useful to review graphs of inbound and outbound traffic from network device interfaces. In Figure 2.3, the graph shows the traffic coming from one device's inbound and outbound interface over a period of time. Graphs like this are created by the NMS at regular intervals and can be used to watch an interface for overuse or underuse.

In Figure 2.3, the X-axis of the graph shows the time of day, while the Y-axis shows the traffic in bits/second coming from the interface. These types of graphs are useful for analyzing short-term traffic patterns because you can see that both your outgoing and incoming traffic shows a slight uptick during the middle part of the day. This can be important in finding the source of a performance slowdown based on the time of day. In this example, depending on the capabilities of your NMS, you might be able to drill down further to see what types of traffic increase during this time period. Maybe users are doing their daily Web surfing during their lunch hour and this extra stress on the network is causing network applications to slow.

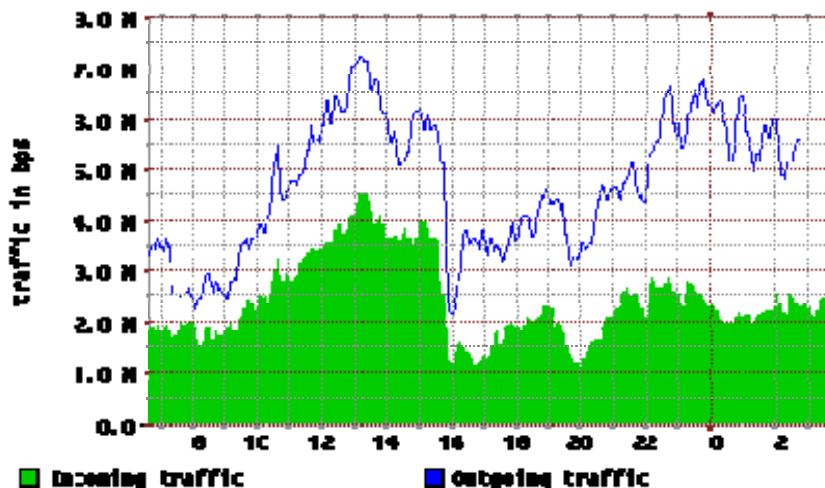


Figure 2.3: Network graphs provide a handy way to identify short-term traffic trends.

- It is important to note that these sorts of traffic graphs are usually good for short-term and medium-term analysis. However, long-term analysis can be difficult due to the impact of non-business hours. When users are only using the network from 8:00am to 5:00pm but the traffic analysis graph monitors throughout the entire 24-hour cycle, compounding errors from those unused time periods can cause averaging effects in long-term graphs. These averaging effects have the tendency to reduce the graphed total network use in long-term graphs.


Performance Correction

As with fault management, the last step in performance management is actually fixing the problem. Within the performance correction step, you use the information given to you by the NMS, the graphs and trending analysis done in concert with the data in your NMS, and your own network intuition to determine that a change is necessary to bring the network back to nominal performance.

That change can take a number of forms. Adding extra interfaces to key network devices can distribute the load and increase performance. Depending on the type of device and load-balancing capabilities, performance can increase linearly with the number of additional interfaces added to the problem.

In some cases, especially those that involve problems over WAN connections, a change to the routing configuration can solve the problem. Often, though, these WAN-based routing configurations involve cooperation with the carrier provider hosting the WAN. In either case, rerouting traffic away from network hot spots to links with lower utilization can improve performance.

Lastly, in some cases in which performance is suffering substantially, the graphing and trending analysis can show that a complete network or network segment re-architecture may be necessary. In this case, movement of clients and servers closer within the network can increase their throughput. New devices with new technologies can increase performance or add compression or optimization capabilities. Conversion or encapsulation of client-to-server traffic within other protocols can improve performance by changing the method of access completely.

 Not all network performance problems occur at the OSI model's layers 1, 2, and 3. Your performance problem may have nothing to do with physical connectivity or problems with TCP or IP routing at all. Some performance problems have to do with higher-level protocols and their tendency for "chattiness."

Databases and their associated application servers are a great example of this. If you separate a database from its application server over the network, you will often see a substantial performance loss. Consider correcting this problem by relocating the servers closer in terms of network proximity or by using other high-level protocols (such as RDP, ICA, SSH, and so on) to encapsulate the user's connection to their data.

Key Measurements in Performance Management

To properly undertake a performance analysis, you must have an agreed-upon yardstick of measurements that can quantitatively describe the qualitative behaviors on the network. This section will discuss five of these key measurements. The first two, bandwidth utilization and network latency, characterize the movement of data across the network from source to destination. The last three describe the state of the network device and its internal resources as that traffic moves in and out of the device.

Bandwidth Utilization

Before talking about bandwidth, it is important to dispel one common fallacy. The concept of bandwidth utilization is quite possibly one of the least-understood measurements in networking. This is the case because bandwidth and the use of bandwidth is actually not what most people really believe it to be. This comment is best explained through a comparison.

Bandwidth is defined as the amount of information that is physically possible to send through a particular media. The only people who can really talk about bandwidth are the physicists in the room because bandwidth is a measurement of the theoretical maximum capability of a particular network medium. Also, Ethernet is a baseband technology, which means each transmission fully utilizes all the available bandwidth. That is, all computers communicate at the transmission speed of the connecting medium.

A much more accurate description for the type of measurement we mean when we say “bandwidth” is “throughput.” Throughput is used to describe the measure of how much actual data could be sent over that media in a unit of time in the real world. Although this differentiation exists, in the real world, bandwidth and throughput are often used interchangeably, which is why it’s important to highlight the difference.



Bandwidth is equivalent to a medium’s theoretical maximum and throughput is equivalent to that medium’s real-world maximum, so you can consider them to be proportionally related.

No matter what the verbiage, when you think of bandwidth, think of it as the width of the pipe down which you’re trying to send data. The utilization of that pipe is the measurement of how full the pipe is. If the utilization of a 100Mbps pipe is 55Mbps, then you’ve got just about half the pipe left through which to send data.

Bandwidth in performance management is important for network architecture and capacity planning. When designing a network, you must ensure that the bandwidth across every node in the thread is capable of handling the load it will be assigned. As the utilization of available bandwidth grows to near 100%, the amount of time it takes to get data through that pipe increases because the data has to wait for the pipe to empty before it can start on its journey.

Network Latency

This concept of time delay in data transmission segues perfectly into the idea of network latency. Bandwidth gets a lot of attention because of the word's heavy use in the consumer market, but it could be argued that network latency more often than not causes the real problem in a network. In many business networks with 1Gbps links, utilization rarely goes above 10%. However, the time it takes for the data to traverse the network can range from less than one millisecond for a LAN connection to hundreds of milliseconds for a satellite connection to another continent.

```
C:\>ping 65.254.250.110

Pinging 65.254.250.110 with 32 bytes of data:

Reply from 65.254.250.110: bytes=32 time=80ms TTL=237
Reply from 65.254.250.110: bytes=32 time=80ms TTL=237
Reply from 65.254.250.110: bytes=32 time=80ms TTL=237
Reply from 65.254.250.110: bytes=32 time=80ms TTL=237

Ping statistics for 65.254.250.110:
    Packets: Sent = 4, Received = 4, Lost = 0 (0% loss),
    Approximate round trip times in milli-seconds:
        Minimum = 80ms, Maximum = 80ms, Average = 80ms

C:\>
```

Listing 2.1: *In Microsoft Windows, a rough estimate of network latency can be obtained by pinging a remote host and reviewing its round-trip time.*

As Listing 2.1 shows, you can use the PING command in most OSs to send a series of packets to a remote host and measure the amount of time it takes for that packet and its associated acknowledgement to complete a round trip back to the originating host.

Network latency is a good measurement of performance because it can be related to a network link's utilization and is much more easily measured. When a link's utilization is very high (for example, the pipe is full), network latency can be similarly very high (for example, data is waiting). One good measurement for identifying the performance on a network is to measure the latency between a client and its application server. Each device and link between devices in the thread between client and server (also known as a "hop") will add to the latency. Your performance baseline should recognize what are acceptable latency measurements, and your NMS should monitor for when that latency goes above acceptable measurements.

Interface Errors and Discards

Interface errors can occur when a problem exists on the network, such as a bad cable, line noise, or a malfunctioning device. Although interface errors are usually used as a metric for detecting faults on a network, their presence on a network device can help deduce an unexplained slowdown in network traffic. When interface errors occur, the traffic that is lost or corrupted in its transition to that interface on the network device needs to be re-requested and retransmitted. In situations in which substantial retransmission occurs, a significant reduction in performance can occur—even to the point of going beyond timeout conditions. Your NMS should monitor for interface errors and alert on situations when interface errors on an interface of a device exceed zero.

Network discards are rarer and are not necessarily always the product of a network problem. Discards occur when the policy of a network device instructs the device to ignore the traffic coming in on that interface. This happens often when device policies or Quality of Service (QoS) policies are enabled on the interface. In these situations, if the traffic is legitimate and you intend for it to be routed, the device should be reconfigured not to discard the traffic. Remember that some traffic likely should be discarded based on the policies assigned to the network device. In any case, similar to errors, NMS monitoring of discards will help identify when this situation occurs.

```
router1#show int
Ethernet0 is up, line protocol is up

[...lines removed...]

 11809 packets input, 933204 bytes, 0 no buffer
Received 8612 broadcasts, 0 runts, 0 giants
 0 input errors, 0 CRC, 0 frame, 0 overrun, 0 ignored, 0 abort
 0 input packets with dribble condition detected
 7737 packets output, 651486 bytes, 0 underruns
 0 output errors, 0 collisions, 1 interface resets, 0 restarts
 0 output buffer failures, 0 output buffers swapped out
```

Listing 2.2: Some of the interface statistics on a Cisco router will provide information about interface errors.

Network Hardware Resource Utilization

Like any computer, a network device is a data processing device. And, as with any data processing device, there are hardware resources onboard a network device that it needs to do its job. If those hardware resources are being overutilized, the network device will not be able to quickly process network traffic.

Statistics on resource use are available for most network devices and your NMS should monitor for their overuse. During nominal conditions, network devices do not typically incur substantial processing requirements except in exceptionally high-use conditions. Thus, if a network resource begins to incur substantial resource use, it is likely that there is a problem on the network. That problem can be related to a device with debug logging enabled at too high a level, an overactive inbound interface, or a security situation involving some hacking attempt.

There are three very important statistics on network devices to monitor in terms of hardware resource utilization: CPU load, memory usage, and buffer usage. The next section will discuss buffer usage; the other two are both important counters because of their direct impact on that device's ability to perform its function. Your NMS should provide the capability to display monitoring information regarding device CPU and memory utilization and alert based on overuse of those resources.

Buffer Usage

Network devices typically incorporate a series of memory blocks to be used when transferring data between the internal components of the device. If an interface needs to send a packet of data to a routing processor, it must first reserve a buffer location to store the data packet. Failures with this buffer request and assignment process can be one of the biggest factors in packet drops, which lead to retransmissions and ultimately network performance loss.

A number of metrics of buffer use can be monitored and managed through the NMS. These metrics include:

- Total buffer number
- Number of permanent buffers
- Number of buffers in free list
- Buffer hits/misses/trims
- Buffer failures

Although buffer tuning is not normally a necessary task with network devices, knowing the status of available buffers and their hit and miss rate can help identify the source of network performance lag. Like the other metrics discussed in this section, your NMS should have the capability to interrogate and report on buffer metrics.

```
router1#show buffers

Buffer elements:
  500 in free list (500 max allowed)
  2370 hits, 0 misses, 0 created

Public buffer pools:
Small buffers, 112 bytes (total 16, permanent 10):
  12 in free list (0 min, 10 max allowed)
  1770 hits, 33 misses, 22 trims, 28 created
  8 failures (0 no memory)
Middle buffers, 600 bytes (total 90, permanent 90):
  90 in free list (10 min, 200 max allowed)
  595 hits, 0 misses, 0 trims, 0 created
  2 failures (0 no memory)
```

Listing 2.3: The Cisco `show buffers` command can provide information about buffer status.

The Business Metrics of Performance Management


A 20% reduction in the response time of a critical network application can potentially lead to a 20% reduction in the productivity of the worker using that application. Understanding the interrelatedness between network application performance and its affect on worker efficiency is critical to business. Conversely, there can be an equal loss in IT worker performance if those workers are constantly tracking down “the network is slow” comments. For a business to get the most out of the production network, a few tried-and-true metrics have been developed that help the technology people in IT relate to the dollars-and-cents people in business. This section will discuss a few of the metrics you can incorporate that leverage the monitoring capabilities of your NMS to provide useful reports to your business leaders.

Availability and SLAs

The concept of availability can be defined as the ability for workers to access the data and applications they need to accomplish their daily tasks. As Chapter 1 discussed, availability metrics are usually measured in whole-system terms: Does the user have access? Is the network up? As most modern networks incorporate some form of redundancy in order to increase availability, business is typically less interested in individual interface outages except when those outages affect that fundamental question.

Service Level Agreements (SLAs) can be created as contracts between IT and the business that outline the performance and availability requirements enforced upon IT by the needs of the business. These contracts establish, among other things, minimum acceptable performance metrics for applications on the network. Creating an SLA means that you and the business have agreed upon set standards for application performance. They also give you ammunition for identifying when hardware purchases are necessary to meet that SLA and for setting the quantitative standard for combating those “the network is slow” comments from your users.

One other major area in which SLAs are used is within purchase decisions of network carrier providers. When going through purchase decisions for WAN connectivity over commodity networks, SLAs with the carrier provider are necessary to ensure your nominal performance of the WAN link. Provider carriers are notorious for holding their customers to the precise verbiage of the SLA contract. This is due to the substantial cost of chargebacks to the customer when the provider's performance breaks SLA guidelines.

 SLAs are contracts. Thus, the verbiage in the SLA is held to a high standard. If you require an SLA from your business or (more importantly) a carrier provider, check the fine-print definitions and reimbursement conditions very carefully.

Because of this, be aware in contract negotiations of those contractual metrics you want to enforce on your carrier provider. Table 2.1 provides a list of network performance metrics that can be defined within an SLA as well as the threshold for breaking that metric. For each line item in the table, a typical SLA will also specify what the cost or amount of the chargeback will be to the customer if the threshold is exceeded.

Metric	Threshold
Availability	99.9% uptime—equivalent to 8.7 hours of downtime per year
Mean Time to Restore	4 hours
Committed Information Rate (CIR)	512Mbps (burstable to 1Gbps)
Latency	< 50ms

Table 2.1: Example SLA metrics with a carrier provider and the associated thresholds.

Bandwidth Monitoring

With some notable exceptions, utilizing WAN connections through network carrier providers rarely means a direct point-to-point connection. Frame Relay connections are a particular type of WAN connection through the carrier provider's network that can bounce over numerous hops while within the provider's network. The provider does not provide a specific connection from one site to another. Instead, they provide a guarantee that the traffic will exit "the cloud" within a predetermined amount of time.

These types of connections are excellent for SMBs and businesses in the mid-market because dedicated point-to-point connections are expensive. By using Frame Relay connections, the connection from your home office to your remote offices goes into your provider's internal "cloud" once it leaves your internal network. You are not responsible for maintaining a direct line from office to office. Through SLAs and other contracts, your network carrier affirms a particular performance for those connections.

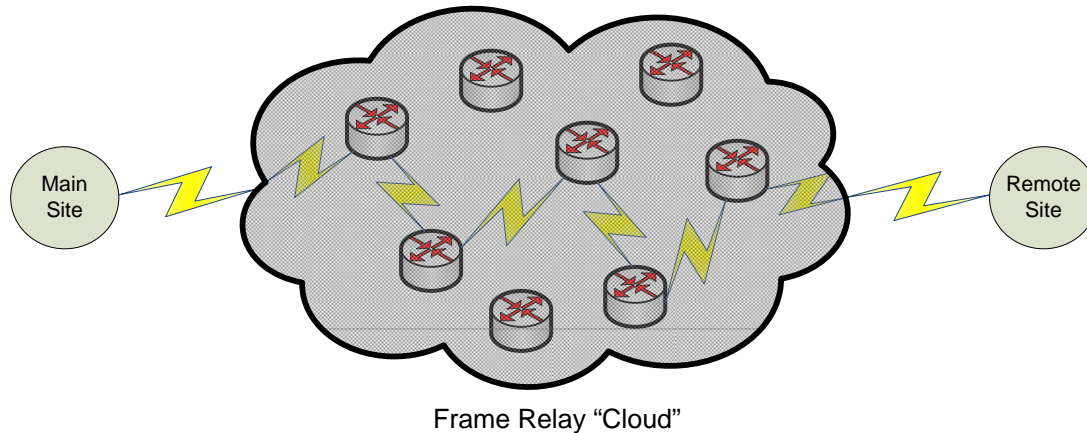


Figure 2.4: Inside the carrier provider's Frame Relay "cloud," your main site traffic can route through any number of hops before it exits at your remote site.

However, it is your money buying that link. Thus, it's often up to you to manage the monitoring of that link's bandwidth and other performance characteristics. In some contracts with carrier providers, the onus of notification is laid on the customer to notify the provider when the link is not performing.

In this case, your NMS can also assist with measuring bandwidth and latency metrics as well as notifying when a link drops. Bandwidth can be approximated by measuring the amount of time taken to move a data structure of a known size from one end of the pipe to another. Latency can be measured by round-trip time for ping packets. With an effective NMS, measurements are available out of the box to measure bandwidth characteristics across Frame Relay connections.

Whenever using carrier providers, consider configuring your NMS to regularly monitor the state of network connections and beyond just up/down notification. As the responsibility can be upon you the customer to notify (and receive the appropriate chargebacks) when the carrier's connection degrades, configure bandwidth monitoring to alert when conditions are out of specifications.

Link Costs

Link costs are a network term that describes the relative network "cost" associated with sending a data packet across that link. Network routing protocols such as Open Shortest Path First (OSPF) utilize link costs to make routing decisions when given redundancy options. Each link in a network is assigned a number whose absolute value is meaningless but whose proportional value in comparison with all other link costs in the network determines how traffic is routed.

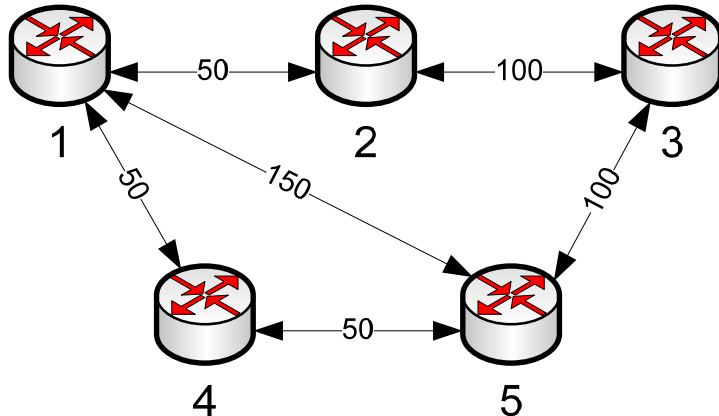


Figure 2.5: Each connection between routers is assigned a link cost. That cost is used to determine the least costly route to move data across the network.

The best way to illustrate this concept is with a picture. In Figure 2.5, for data to move from Router 1 to Router 3, it has three possible paths:

- Option 1: Router 1 → Router 2 → Router 3 = 150
- Option 2: Router 1 → Router 4 → Router 5 → Router 3 = 200
- Option 3: Router 1 → Router 5 → Router 3 = 250

Option 1 has a link cost of 150 added across the two hops from Router 1 to Router 2 to Router 3. This is the lowest of the possible routes the packet can take without retracing its steps across a hop. Thus, the traffic will default to Option 1.

Link costs are useful to monitor when redundancy options are added to a network. As you can see from Figure 2.5, if the connection from Router 1 to Router 2 goes down, it is still possible to send traffic to Router 3, hopping through an alternative path. Although Option 2 involves more hops, it will become the backup path because its link cost is lower than Option 3.

Link costs are important to monitor and maintain as business metrics because the link costs should be relevant to business application performance during non-optimal conditions. If your routing is configured such that a failure causes application performance to go beyond acceptable thresholds, a network rearchitecture or rerouting analysis may be necessary.

Traffic Management and Prioritization

The last business metrics that should be monitored for performance management reasons are traffic and prioritization. The Internet is a great place for research and for communicating with others in your industry, but it's also a great place for streaming music and video and overactive Web surfing. All this activity can impact the performance of your network connection to the outside world.

As noted during the performance graphing discussion, it is reasonable to assume that some Internet browsing by employees will occur during the lunch hour (unless you have a highly restrictive policy against employee surfing!). However, overuse of your corporate network for employee Internet browsing can have a detrimental effect on your network performance.

The concepts of traffic management and prioritization go hand in hand. Traffic management is the idea of implementing a policy-based approach to network traffic, granting or denying that traffic based on centralized rules. Traffic prioritization takes that one step further by allowing network devices to prioritize important traffic (for example, critical database to critical application server) over that of non-important traffic.

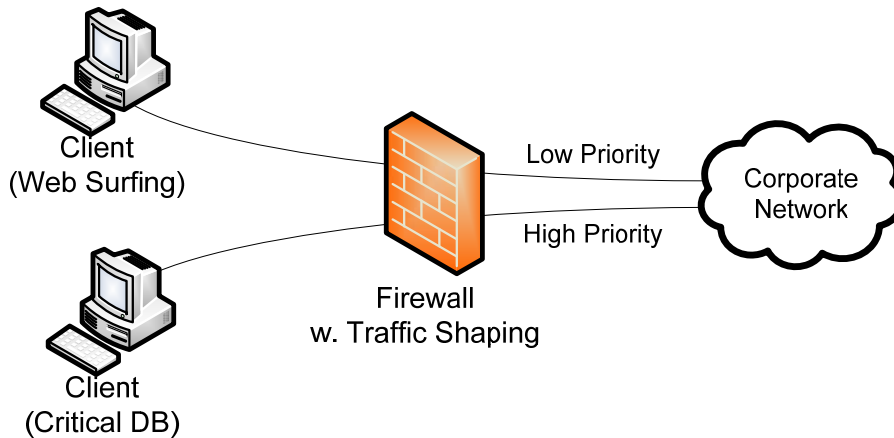


Figure 2.6: Firewalls and other network devices that incorporate traffic shaping can prioritize critical traffic over non-critical traffic within a network.

Both mechanisms require some form of management tool that enables policy-based management and traffic shaping to monitor, validate, and route traffic correctly. But in networks constrained by budgetary restrictions, enabling traffic management and prioritization can proactively reduce the productivity loss associated with misuse of available network resources.

Additional Tools for Managing Performance

Throughout, this chapter has discussed a number of tools—both procedural and technical—that enable you the network administrator to better manage the performance of your network. This section will discuss three additional tools that help in specific situations: Simulating loads, analyzing traffic patterns, and monitoring wireless communications.

Traffic-Generation Tools

In some situations, the only way to effectively identify when network performance will go critical is to simulate load on that network link. Traffic-generation tools are used for just that purpose. They generate an abundance of traffic to route over a configured connection. This administrator-configurable traffic can be adjusted to simulate increasing load over a connection. While the load is manipulated, application response time is judged during each phase in the traffic loading.

As with any network traffic, these tools must have a host on the receiving end configured to accept the incoming traffic. Typically, traffic from these tools is sent to destination port `udp/9`, which is configured on UNIX systems as a “discard port.” All traffic routed to `udp/9` is read by the destination network interface and immediately discarded. This concept of a discard port allows for the successful receipt of traffic without overwhelming the destination system with the volume of traffic being generated by the source.

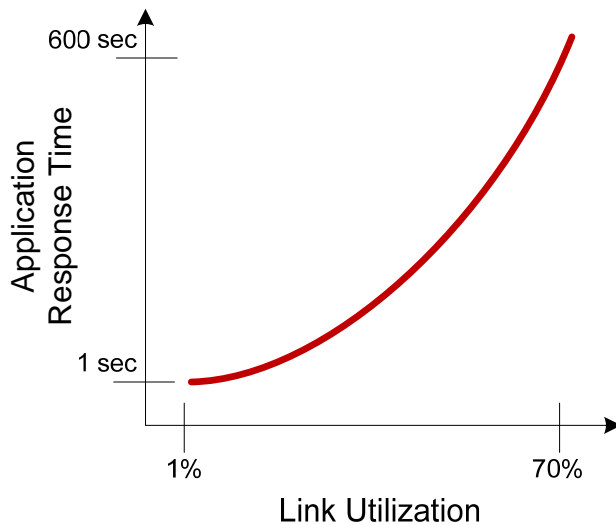



Figure 2.7: Increasing link utilization can have a negative effect on application response time. Traffic-generation tools are used to simulate this link utilization.

Traffic-generation tools are often components of the network toolkits that accompany NMS software packages. To successfully simulate a number of conditions, they typically include the following configurable characteristics:

- Target name or IP address
- Port number, UDP or TCP
- Packet size
- Percentage of circuit bandwidth to consume

 Because traffic-generation tools have the capability to generate substantial amounts of traffic, they can be misused as tools to generate Denial of Service (DoS) attacks.

Traffic-Analysis Tools

Traffic-analysis tools are those that allow for collection of performance data and graphical analysis of that data across multiple network devices. These tools are used during performance baselining and performance reporting activities to gather an understanding of the underlying traffic on the network, its type and time of day, and its source and destination.

Effective traffic-analysis tools will be highly graphically oriented to provide the administrator with charts and tables representing a global view of the network. These charts and tables typically have drill-down capabilities to provide the administrator with more detailed information about a particular traffic flow. As the types and network protocols involved with network traffic are many and complicated, a good traffic-analysis tool will also be able to identify within each traffic flow the protocol type and graphically display those types for analysis.

There are generally two types of mechanisms for analyzing traffic across a particular network connection. The first involves the incorporation of a network probe device that is installed in-line between two network devices. The network probe monitors the traffic between the two devices and reports what it sees to the NMS. The second uses a technology developed by Cisco called NetFlow. Built-in to many network devices, the Cisco NetFlow technology eliminates the need to incorporate an in-line device. Instead, the traffic is monitored at the interface of each of the NetFlow-enabled devices. Using NetFlow reduces the administrative overhead of installing, managing, and ultimately removing network probes.

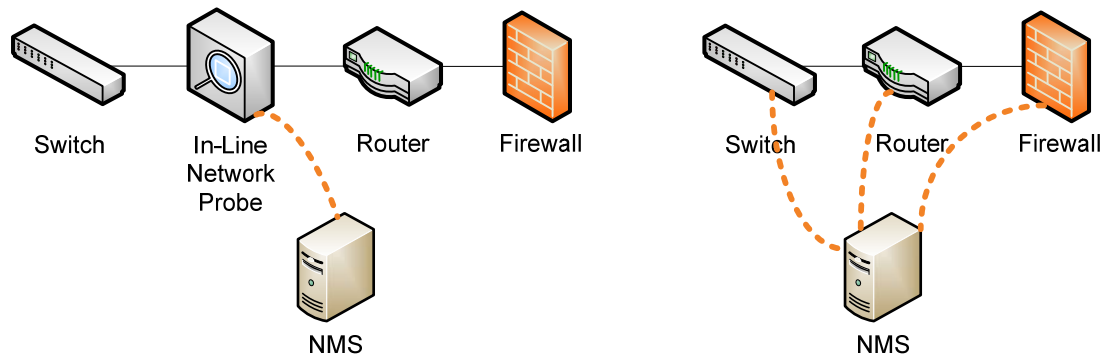


Figure 2.8: On the left, an in-line network probe sends SNMP performance data to an NMS. On the right, an NMS communicates with Cisco NetFlow-enabled devices to probe the internal performance counters on each device.

No matter which type is used, traffic-analysis tools complement the aforementioned traffic-prioritization tools that identify and prioritize traffic types based on administrator-set policies. These tools enhance mere analysis by allowing administrators to actively fix overactive protocols on their networks. Characteristics that can be analyzed by these types of tools include:

- Global view of consumption
- Consumption over time
- Consumption by hour of day
- Application consumption patterning
- Type of traffic
- Traffic data archival
- Historical analysis
- External traffic sourcing

Wireless Performance Tools

Wireless networks have their own special needs—from the additional security needs that ensure hackers can neither see nor access internal networks to the location-based complexities of dealing with a network intended to operate through walls. These special needs require special administrative tools. Outside the typical feature sets associated with NMS' and traffic analysis tools, wireless performance tools typically provide additional features that enable access point signal strength metering, mapping of the network to identify “hot” and “cold” spots, and identification and analysis of clients and sessions.

Arguably, the most difficult part of wireless administration is the management of the Wireless Access Points (WAPs) in a network. Because wireless is designed so that users can roam from any WAP to any other WAP, a good wireless performance tool will provide information about dropped network packets, roaming patterns, and active devices. As devices move throughout the network, the administrator will likely want to identify clients, their attached WAP, and their signal strength to that WAP. Doing so will assist the administrator with identifying where building features may be interfering with the wireless signal. Wireless typically operates at a much lower bandwidth than wired connections, so good tools also can show the administrator where bandwidth contention is occurring between clients on the same WAP.

Performance Affects Business

As discussed earlier in this chapter, a 20% reduction in the response time of a critical application can potentially lead to a 20% reduction in workers' productivity. That 20% can mean the difference between an agile, successful business and one that is unable to keep up with the needs of its customers. As you can see throughout this chapter, there are numerous tools available that can provide a quantitative solution to what has historically been a qualitative problem—“the network is slow.” With these tools in place, you can proactively identify the slow spots on the network and begin implementing solutions before that call occurs.

The next chapter moves away from problem management and focuses specifically on the tenants of configuration management. Chapter 3 talks about how you can employ an NMS that can ensure a consistent configuration across your network devices. This consistent configuration will ensure your adherence to compliance regulations as well as ensure a consistent and repeatable configuration to all the devices on your network.

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>.