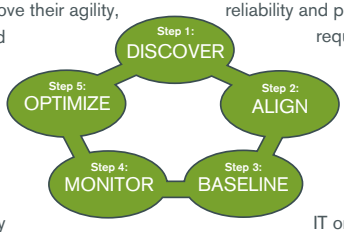
Application and Network Performance Analysis

A Five-Step Process for IT Professionals

NETWORK GENERAL | Know the Network™

Introduction

The enterprise IT organization has changed in recent years as a result of the economic recession and subsequent downsizing requirements. Today, IT executives must do more with less, and this trend is expected to continue. In fact, many IT organizations are fundamentally changing the way they are structured and the way they operate. Rather than focusing on cost-cutting, they are looking for ways to improve their agility, efficiency and strategic value to the business.



Agility and efficiency are certainly compelling goals, but how can you make meaningful progress in achieving them? One way is to implement a structured, multi-step process which, when consistently applied and effectively linked together, can lead to sustainable positive outcomes in quality, efficiency, productivity, and customer satisfaction.

This type of process doesn't need to be complex to be effective. On the contrary, if a process methodology is simple, it will help ensure that it is well understood, easily adopted, and successfully executed by all participants in the process. Within IT operations, there are many areas where processes can be (and have been) successfully deployed. This document focuses on just one specific aspect of IT management – the performance of enterprise application and network resources. More specifically, it focuses on the steps involved in ensuring that IT professionals are armed with the necessary information to ensure critical application and network resources are delivering the reliability and performance required to drive the business forward. In addition, this process can also help IT organizations extract maximum performance and value from existing staff and budget resources.

This process methodology was created based on the collective experience of Network General. For nearly two decades the company's service and support professionals help resolve IT performance problems for thousands of customers worldwide. However, there is no mention of specific products from Network General or any other vendor in this guide. The discussion here covers only generic functions and capabilities which must be incorporated into the process to gain the maximum benefit and outcome.

Step 1: Discover

An effective process begins by understanding – or in many cases "discovering" – the complete application and network environment. From a network perspective, that means first taking an inventory of LAN and WAN connections across the entire distributed enterprise. This may seem like either an obvious or unnecessary step, but it is useful to construct a "bandwidth map" that begins in the corporate data center and extends to the desktop of every remote user.

This step in the process will reveal variations in available bandwidth, which may result in variations in network performance. For example, it is likely that bandwidth availability and requirements vary at the LAN/WAN boundary, in some remote offices, and in dense server farms. This information will provide a useful foundation for subsequent steps in the process.

But as important as a network bandwidth map is, it offers limited value – and could actually be misleading – without also discovering all of the applications that are consuming available bandwidth. This is arguably the most important step in the entire methodology outlined here. The only reason networks exist is to support the applications that companies depend on to help run the business. According to analyst estimates, businesses worldwide collectively spend more than $40 billion on enterprise application software each year, and that does not include legacy or custom-developed software. Clearly, there is an enormous investment in, and dependence on, software to run the business.

> What's needed are tools that identify all the applications traversing the network, providing details on how those applications are being used.

Unfortunately, business applications are not the only thing consuming bandwidth. User accessibility to the Internet also means accessibility to a broad range of applications, utilities, and files that are downloaded to user desktops and shared across the corporate network. Some of these may be useful but most of them are not – including music files, gaming, videos, P2P traffic, and other recreational applications. All unsanctioned applications and traffic can contribute to network and business performance problems,

and cause massive administration headaches for IT professionals. Even with the best IT policies in place, this remains a chronic problem in many organizations.

What's needed in the initial discovery phase are tools that identify all the applications traversing the network, providing details on how those applications are being used. That means identifying the unique signature of specific applications in order to distinguish between Siebel and Oracle, for example. In addition, it is necessary to understand how applications map to specific client/server transactions that occur each day across the enterprise network.

For example, who is the user? Where is the user located? What application is involved in the transaction? What servers are involved in the transaction? Where do the servers reside? What network links support the transaction? What are the application and network protocols, IP addresses, port numbers, and MAC addresses associated with the transactions? What are the bandwidth and response time requirements to ensure this application performs well?

In short, the discovery phase needs to provide a *depth* (Layer 2-7 profile of transaction data) and *breadth* (end-to-end, from remote client to corporate server) of information necessary to gain a holistic view of the application and network landscape across the extended enterprise. Think of it as an architectural blueprint of your application and network infrastructure. The next challenge will be to optimize the blueprint as necessary to support the requirements and priorities of the business.

Step 2: Align

One of the primary objectives of every CIO is the "alignment of IT with business." But what does that really mean? In the context of this focused discussion on application and network performance analysis, alignment means that network resources are properly tuned to ensure maximum performance and reliability of the critical applications that power the business.

This step in the process is somewhat unique because it focuses less on technology and more on the business. During this phase, it is essential to invest sufficient time with the business leaders in the company, getting answers to several important questions.

What are the primary business objectives in the current fiscal year? Who are the major business stakeholders responsible for achieving the objectives? Which applications support the processes necessary to reach these business goals? Will there be specific times when application availability and performance are most critical?

This step in the process is somewhat unique because it focuses less on technology and more on the business.

What other users are involved in these processes and where are they located in the organization?

The plethora of information on application and network resources captured during the discovery phase becomes invaluable at this point. Armed with that discovery data as well as an understanding of business priorities and processes, it is now possible to determine strategies for application and network resource optimization to help improve the alignment of IT with business.

Let's say, for example, that we're dealing with a Fortune 500 company where ERP is the primary back office application supporting a variety of financial processes across the extended enterprise. And let's assume that VoIP is also deployed in several regional offices, connecting users at local and remote sites across the WAN. Both of these applications may involve hundreds of users, and both have certain stringent bandwidth and response-time requirements. If their performance requirements are not met, the performance of individual employees and the business will suffer as well.

In this scenario, it would be helpful to identify the communication paths between all the clients and servers involved in the voice and data transactions. The discovery data would reveal where the users are located, what bandwidth they have available, where the servers are located, and information on any other non-critical applications that may compete for bandwidth and cause congestion.

But knowing this information is not enough. From an IT management standpoint, what is needed is a tool that could, for example, allow IT professionals to create multiple, business-critical transaction "profiles." Each profile could consist of a specific user, application, network connection, and server resource involved in a specific VoIP or ERP transaction.

By associating these different elements into a composite transaction profile, IT professionals could set performance thresholds for each business-critical process, then monitor the performance of the profiles very closely to ensure application and network service level objectives are being met. In fact, each transaction profile could have its own unique service level management objectives. With this granular level of monitoring and control, alignment becomes a more attainable objective.

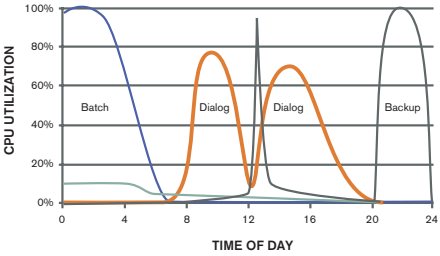
But how should appropriate performance thresholds be set for the various transaction profiles? The answer to that question begins by discussing the next step in the process – establishing a performance baseline.

Step 3: Baseline

A baseline represents the normal, acceptable performance level for application and network resources. Separate baselines should be established for the critical variables captured during the discovery phase, including application response time, server processing, network link utilization, and more.

Establishing accurate baselines is an essential part of effective IT resource management, but this process is where art and science have typically come together. To establish a baseline, IT professionals must examine all the performance data (the science), then take a swag at determining the normal, acceptable performance level for a particular IT resource or operation (the art). The result is often a graph showing the performance data of a resource over time, with a horizontal line cutting across to show the baseline of acceptable performance.

Once baselines are established, thresholds can be set at some point above or below the baseline to represent the point at which performance is no longer acceptable. If performance deviates to the point that it exceeds a performance threshold, an alert would typically be sent to an IT manager with relevant information so that action can be taken even before any disruption is noticed by users on the network.



This process sounds quite reasonable, but real world conditions usually render it ineffective. Why? Because real world operations do not occur along a straight line. This is evident from the graph above, which shows server CPU utilization captured from an ERP on-line transaction processing environment during a 24-hour period. The burstiness of server utilization is caused by the reality of varying, often unpredictable, usage throughout the day.

Imagine trying to establish a single horizontal baseline for a resource like this. Threshold alerts would be triggered throughout the day, sending IT administrators into a frenzy. As a result, IT administration costs would go up, IT productivity would go down, and baselines would provide no real value to anyone. This scenario occurs all too often in real world networks simply because there are typically wide swings in transactions and resource utilization throughout the enterprise each and every day.

IT professionals sometimes try to solve this problem by over-provisioning a resource, then setting the baseline at the high end of the range. The greatest benefit of this approach is that fewer performance alerts are sent to the IT staff. But this strategy also increases capital costs without necessarily ensuring alignment of IT resources with business processes and priorities.

The best solution to this problem is to ensure that performance baselines and their corresponding thresholds reflect the reality of varying application and network performance levels during the course of any business day. As noted earlier, that's not possible with the traditional approach of establishing static baselines and thresholds. But what if these metrics were "dynamic" in nature?

For example, in the earlier graph, dynamic baselines would more likely appear as undulations that shift with normal, expected resource usage throughout the day. Likewise, dynamic thresholds would follow the same path, at specific points above and below the changing baseline. As a result, if performance levels went outside the boundaries of the dynamic baseline range, IT alerts would be very meaningful and would be taken far more seriously.

The best solution to this problem is to ensure that performance baselines and their corresponding thresholds reflect the reality of varying application and network performance levels during the course of any business day.

Now consider the possibilities of applying dynamic baselines and thresholds to the transaction profiles discussed earlier. In such cases, IT professionals could create transaction profiles for all ERP users, where each profile included the client, network links, application, and server resources. Baselines and

corresponding alert thresholds could be adjusted based on each user's typical patterns during the day. That way, if the parameters of a transaction profile exceeded threshold boundaries, the subsequent alert would more likely reflect a valid performance problem. Just as important, application and network resources would be more aligned with the need and realities of the business.

Once dynamic baselines and thresholds have been applied to either specific critical applications or to individual transaction profiles, the next step is to closely monitor application and network performance.

Step 4: Monitor

This is a vitally important phase where IT professionals should be actively involved in measuring and analyzing data on application and network performance. After all the preparations that are made in the first three steps, this is where the reality of day-to-day enterprise network computing comes into play for the first time. It is, therefore, essential to learn as much as possible about what's working, what's not, and what will need to be adjusted in the next step of the process.

To get started, the enterprise network should be properly instrumented with distributed monitoring tools that are deployed at critical points throughout the network. These tools should ideally gather a rich set of information on network, application, and server resources and operations, and reveal potential performance problems associated with network latency, transaction response time, resource availability, and much more. If desired, a subset of this data can then be forwarded to a central management console for action by the IT staff. It is advisable for IT organizations to instrument their networks with both real-time monitoring and long-term historical monitoring tools.

It is advisable for IT organizations to instrument their networks with both real-time monitoring and long-term historical monitoring tools.

Real-time monitoring is important because it provides a "live" view into what's happening on the network. This is especially important for tracking performance-sensitive applications such as VoIP, and for safeguarding the network from potentially malicious traffic. And, when the network has been properly instrumented for remote monitoring, IT professionals can gain immediate visibility into these types of problems no matter where they may occur across the distributed, global enterprise.

Long-term historical monitoring is a relatively new concept that also offers tremendous value in application and network performance analysis. Let's assume, for example, that an ERP user is experiencing very slow application response time – not consistently, but randomly over a two-week period. Each occurrence of the problem lasts only 5-10 minutes, but transactions are delayed and productivity is impacted

as a result. By the time IT professionals get a chance to identify the problem, it is gone.

With a historical monitoring tool in place, data would be automatically collected and stored over several weeks. Thus, in our example, when the ERP response time problem is reported by the end user or from an automatic alert, IT professionals can review and analyze the historical data that has been collected, identify the random occurrences, and discover, for example, that unscheduled batch processing jobs from the same office are randomly impacting ERP performance by causing network congestion and consuming server CPU time. Without that historical perspective, it would be far more difficult to isolate and resolve the problem.

Both real-time and historical monitoring can – and should – be applied to the transaction profiles discussed earlier because these profiles represent operations and resources that have been deemed "business-critical." Although dynamic baselines would have been established for these profiles, it is still important to closely monitor each profile and its individual components to determine whether the baselines and thresholds are properly configured. In addition to monitoring profiles, the short and long term data captured during this phase is also invaluable in assessing whether application and network resources are properly tuned and IT and business alignment objectives are being met.

Despite close adherence to the first four steps in this process, the unpredictable realities of enterprise computing make it necessary to address one more very important step – the need to continually optimize.

Step 5: Optimize

Change is a constant within the enterprise – new offices, new users, new connections, new applications, etc. And all of these changes can have a profound impact on the operational state of the current network. Therefore, despite the most careful preparation, there will always be a need to repair, tune, and continually optimize the performance of application and network resources.

Every network is unique, as are the problems associated with them. A good starting point is to first determine whether the root cause of a performance problem is related to the network or to a specific application. Knowing this ends much of the finger pointing that often occurs between application and network groups in IT and, more important, equips IT professionals with the information necessary to narrow the focus and move toward a speedy resolution of the problem. This requires tools capable of providing in-depth analysis of application and network transactions based on data captured from all relevent layers of the OSI model.

A good starting point is to determine whether the root cause of a problem is related to the network or to a specific application.

This level of thorough root-cause analysis often reveals a variety of problems, including misconfigured network routers, malicious traffic inside the firewall, inadequate WAN bandwidth, unsanctioned music downloads from the Internet, poor load balancing in the server farm, congestion from unexpected FTP transfers, excessively chatty applications, and more.

Each of these problems may require a different course of action to solve the problem (it is beyond the scope of this paper to address all the myriad possibilities), but the important point is that a rich set of application and network data is essential to solving problems and optimizing performance.

It's also important to emphasize that optimization is more than just a reactive response to existing problems. It is equally important to take a proactive stance and look at ways to optimize before a problem occurs. For example, IT professionals should determine whether anything can be done to further optimize the performance of the transaction profiles created

earlier in the process. Because each profile can consist of multiple elements, it is imperative to examine the application and network resources that make up the profile and proactively analyze (and optimize) their performance – before a problem occurs. This critical management task helps ensure that service level commitments are met, and improves proper alignment between IT resources and business processes and priorities.

Finally, the optimization phase should also look at ways to leverage current application and network performance data to make smarter decisions about network design, application rollout, and capacity planning moving forward. This is all much easier to accomplish if analysis is based on a very rich set of underlying data. Choosing the proper network instrumentation and analysis tools is, therefore, a critical decision.

These five steps, implemented as a continuous process flow, will enable IT professionals to improve staff productivity, control IT costs, and deliver more strategic value to the business.

The Next Step

The five steps outlined here are meant to provide a field-proven framework for effectively managing the performance of application and network resources and ensuring they are supporting the business objectives of "agility and efficiency" mentioned at the outset. That said, this methodology – which is really a continuous, iterative process – intentionally leaves many details and questions unanswered. Why? Because the methodology, as well as the underlying management tools and technologies, need to be adapted and customized for the most unique IT environment in the world – yours.

So what's the next step? We invite you to contact Network General to learn more about how this process can be effectively adapted and applied to your IT organization. Network General offers a unique set of consultative services – as well as a broad set of application and network performance analysis solutions – that will help you better understand your application and network environment today, and better prepare your IT infrastructure for tomorrow. Since 1986, Network General has trained and equipped more than 100,000 IT professionals on application and network performance analysis – more than any other IT management vendor in the world.

We invite you to learn more.
Please visit **www.networkgeneral.com**,
or call us at **1-800-764-3337** or **1-972-713-4300**.

About Network General

Network General™ Corporation is the pioneer and global leader in enterprise application and network performance analysis. The Network General portfolio consists of innovative software solutions that allow IT professionals to identify, monitor, measure, manage and resolve application and network performance problems. Supporting the widest range of network topologies in the industry, Network General delivers breakthrough technology solutions that focus on the intersection of the application and the network for comprehensive real-time monitoring, analysis, and reporting across the business enterprise. Network General is a privately-held company headquartered in San Jose, California and has offices in 23 countries around the world.

Network General is built on an 18-year history of industry-leading, high-quality products and services that enable application and network performance analysis. Since 1986, the company has been offering the most intelligent network performance and analysis solutions in the industry, based-on a broad set of patents and an unmatched knowledge of network protocols and application behavior issues. Network General is the trusted advisor to thousands of customers worldwide.

NETWORK™ GENERAL