

EXCLUSIVE SUBSCRIPTION OFFER  
VISIT [WWW.PERFORMANCE-EDGE-JOURNAL.COM](http://WWW.PERFORMANCE-EDGE-JOURNAL.COM)

NETWORK PERFORMANCE  
MANAGEMENT SOLUTIONS  
FOR THE ENTERPRISE

VOL. 3

Fall 2008

[www.performance-edge-journal.com](http://www.performance-edge-journal.com)

# PERFORMANCE EDGE JOURNAL

perspectives:

Rethinking "Safe" Choices: Are Incumbent Systems Really Worth the Cost?

best practices:

VoIP: Do You See What I'm Saying?

tech briefs:

Detecting Threats to Optimal Network Performance

resources:

2008-2009 Calendar of Recreational Network Traffic Madness

**survey:**

Next-Generation Network Operations Survey Results and Analysis - 2

**perspectives:**

Rethinking "Safe" Choices: Are Incumbent Systems Really Worth the Cost? - 9

**best practices:**

VoIP: Do You See What I'm Saying? - 12

**case study:**

Reality IT: A look inside OSF HealthCare - 19

**tech brief:**

Detecting Threats to Optimal Network Performance - 24

**perspectives:**

Technology Issues in This Election Year - 28

**best practices:**

2008 Application Delivery Handbook Excerpt - 34

**tech briefs:**

Understanding and Monitoring Echo Cancellation for Optimal VoIP Performance - 38

Managing the Performance of Financial Trading Applications - 45

**resources:**

2008-2009 Calendar of Recreational Network Traffic Madness - 51

**Editorial Board**

editor in chief:

Brian Boyko

editorial director:

Patrick Ancipink

associate editor:

Jordan Weiss

art director:

Ginger McBride

contributing writers:

Ben Erwin

Jeffrey Hicks

Jim McQuaid

Dr. Jim Metzler

Brad Webster

Performance Edge Journal is published semi-annually by NetQoS, Inc. and distributed free of charge to qualified networking professionals.

Please send direct address corrections and other correspondence to:  
info@performance-edge-journal.com.

Performance Edge Journal copyright 2006-2008 © NetQoS, Inc. All rights reserved. Printed in the USA.

No part of this publication may be reproduced in any form, or by any means, without prior written permission from NetQoS, Inc.

This publication is distributed on an "as-is" basis, without warranty of any kind either express or implied, including but not limited to the implied warranties of merchantability, fitness for a particular purpose, or noninfringement. This publication could contain technical inaccuracies or typographical errors. Later issues may modify or update information provided in this issue. Neither the publisher nor any contributor shall have any liability to any person for any loss or damage caused directly or indirectly by the information contained herein.

**on the web**

questions or comments:

info@performance-edge-journal.com or

visit: www.performance-edge-journal.com.

# LETTER

## FROM THE EDITOR



Brian Boyko

As I speak with IT professionals, two themes seem to reverberate throughout every IT organization: 1) businesses today are finding themselves having to adapt to an ever-increasing reliance on their enterprise networks, and 2) they are focusing more on where they are falling short because "good enough" can seriously interrupt or negatively impact business.

This issue of Performance Edge Journal is devoted to the diverse networking and application delivery responsibilities that today's network professionals must tackle. Industry expert, Dr. Jim Metzler, contributes recent research, presenting some very interesting expectations and realities of today's network operations center (NOC) and how it's not just about monitoring network availability anymore. Luckily, Dr. Metzler provides some relief and guidance through his views on developing and implementing the "next-generation" network operations center.

A related, perennial issue that all IT organizations face is finding the risk-reward balance when considering upgrades to infrastructure and management components. When is it prudent to stick with what you have? And

when does it make sense to invest in more modern technology that may help you progress to the next level of operational maturity? In this issue, we take a look at the dangers of sticking with incumbent technology.

In this U.S. Presidential election year, we find ourselves contemplating the political forces that impact business and the use of information technology. So, we have considered some of the political issues that will impact enterprise IT over the next four years, focusing on eight of them in particular.

Bringing much of this together, we present a case study of OSF HealthCare and its efforts to modernize medical systems over a large, multi-state, wide area network. We also examine specific technology issues,

such as echo and echo cancellation in VoIP solutions, and anomaly detection software that provides early warning of threats to optimal network performance.

Finally, we have a monthly calendar of events through mid-2009 to help you better plan for those periods when your network may be impacted by unusually high volumes of recreational traffic.

Please share your impressions and ideas with us, any suggestions you may have for future content, as well as any questions you would like answered. Write to us at [info@performance-edge-journal.com](mailto:info@performance-edge-journal.com) to help us understand your concerns in your places of work. We encourage you to pass this journal around to colleagues and associates, and get them to sign up for their own free subscriptions.

Thanks for continuing to read Performance Edge Journal. We look forward to hearing from you.

A handwritten signature in black ink that reads "Brian R. Boyko".


-Brian Boyko, Editor in Chief  
[info@performance-edge-journal.com](mailto:info@performance-edge-journal.com)

# SURVEY

Dr. Jim Metzler

NEXT-GENERATION NETWORK OPERATIONS SURVEY RESULTS ANALYSIS:  
A FOCUS ON APPLICATION DELIVERY IS REDEFINING THE NOC





Jim Metzler's recently completed research survey and in-depth interviews detailing the concerns of IT organizations and the structure of the Network Operations Center provided some interesting insights into the way business is adapting to an ever-increasing reliance on enterprise networks - and where it is falling short.

Over the years, computer and communication networks have grown in size, scope, and complexity. Since the early days of DARPA, the Internet's predecessor—almost 50 years ago!—as much research and development has gone into the network itself as into the research it supported, resulting in a technology that now spans the entire globe. We have become dependent on that technology in ways we couldn't have imagined even a decade ago because of that dependency. The need to manage networks has become more and more important to everyday work and life.

The majority of IT organizations are under considerable pressure to evolve to a "next-generation" Network Operations Center (NOC). From a survey conducted of 176 IT professionals, over a quarter of NOCs are perceived as not meeting their organization's current needs. In order to fulfill the current and emerging requirements of today's enterprises, IT staff in the NOC are being driven to do a better job of managing application performance, to implement more effective IT processes, and to troubleshoot performance problems faster.

While the survey results confirmed the conventional wisdom that a NOC is often stove-piped and reactionary, the results disputed the common belief that NOC personnel are focused largely on monitoring in general and that they spend the majority of their time on networking in particular. The survey results highlighted the fact that the inability of the NOC to identify issues before the user does hurts the overall credibility of the IT organization, and that the role of the NOC is often not well understood - even within the IT organization.

The survey results also showed that while the vast majority of NOCs are undergoing significant change, not all NOCs are starting at the same place in terms of the functionality that they currently provide. In addition, IT organizations do not have a common vision of the structure and functionality of the next-generation NOC. To be able to plan for the evolution of their company's NOC in this demanding yet uncertain environment, network professionals need an awareness of what their peers are doing to address the challenges they are facing, as well as an understanding of how well their efforts are succeeding.

## Responding to Trouble

According to survey respondents, the vast majority of organizations have at least a simple escalation process in place for problem response. Ninety-one percent of survey respondents indicated that their organization has a help desk that assists end users, and 80% agreed that the help desk does a good job of routing issues that it cannot resolve to the group that can best handle them.

It should be noted that of the latter group of respondents (those agreeing that the help desk is routing issues accurately), 81% stated that the help desk typically routes issues that it cannot resolve to the NOC. This response is not surprising: three-quarters of the survey respondents also indicated that the network is generally assumed to be the source of application response-time degradation. But in many instances, the NOC also supports a far broader set of functions than just networking.

Cultural reasons may also explain why the help desk typically routes issues that it cannot resolve to the NOC. For example, a manufacturing analyst who was interviewed stated that users tended to contact the NOC for various IT problems because "we have always had the tools to identify the cause of the problems."



“ One of the most important political drivers affecting the evolution of the NOC is the perception of the NOC's current function and value. The perception from inside and outside the NOC can be very different. ”

A manufacturing and security manager stated that as recently as a year ago, his organization had a very defensive approach to operations, focusing on showing that the network was not the source of the current problem. His current motto is "I don't care what the problem is; we are all going to get involved in fixing it." When asked if his motto was widely accepted within the organization, he replied, "Some of the mentality is changing, but this is still not the norm."

One of the clear results of the survey is that the NOC is typically involved in more than monitoring activities. Most NOCs also get involved in problem resolution.

### What about ITIL?

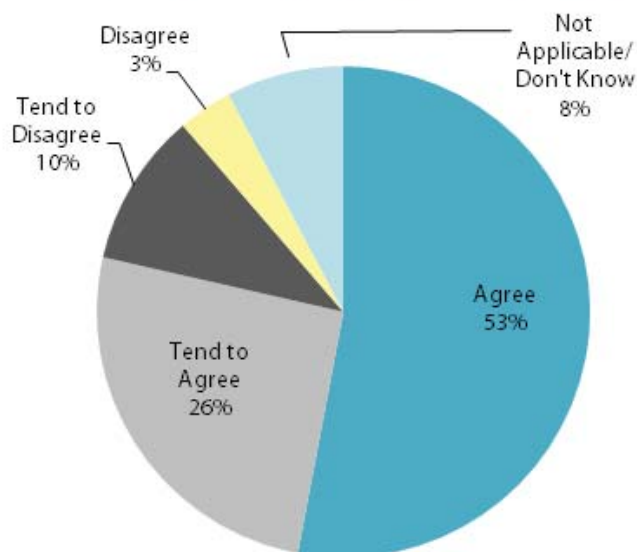
Industry-wide, IT professionals have been involved in significant discussions over the last few years about using a framework such as ITIL (IT Infrastructure Library) to improve network management practices. The majority of survey respondents (62%) indicated that their organizations do have a process like ITIL in place. Of those respondents who said they did not, a similar percentage (63%) said they believed that their organization would put such a process in place within the next 12 months. This response indicates the emphasis being placed within the NOC to improve its processes.

### How Is the NOC Perceived?

One of the most important political drivers affecting the evolution of the NOC is the perception of the NOC's current function and value. The perception from inside and outside the NOC can be very different. This disconnect can have a big impact on NOC staffing and funding - both of which heavily influence the NOC's ability to implement change.

Three-quarters of the survey respondents said they believed that NOC personnel are aware of the key infrastructure components that support the company's critical applications. As such awareness is the basis for aligning IT operations with key business drivers, this response indicates that most respondents view the NOC as being positioned to align itself with the business.

**Our NOC personnel not only identify problems, but are also involved in problem resolution.**



## Survey Results: Do you agree with the statement

Our senior IT management believes that...	Agree/Tend To Agree	Disagree/Tend to Disagree
...the NOC provides value to our organization.	90.7%	9.3%
...the NOC is a strategic function of IT.	87.9%	12.1%
...the NOC is capable of resolving problems in an effective manner.	82.4%	17.6%
...the NOC will be able to meet the organization's requirements 12 months from now.	81.4%	18.6%
...the NOC works efficiently.	80.6%	19.4%
...the NOC meets the organization's current needs.	71.9%	28.1%

Only a small majority of survey respondents (58%) said they believed that the role of the NOC was understood by the entire IT organization. It was heartening, but somewhat surprising, to see that slightly more (63%) of the respondents believed that business managers do understand the role of the NOC. The question of NOC identity was one instance in which there was no significant difference in the responses based on job function.

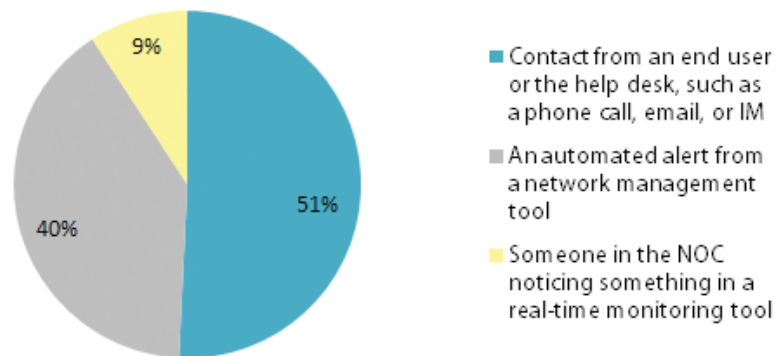
A CIO who was interviewed said a lot of business leaders see "IT" as one big mass, and that this lack of understanding of IT can cause problems. He said, "When times are good, the business managers forgive how much we spend on IT. However, when business is going through a down cycle they ask, 'Why are we paying so much for IT?'"

### What Does the NOC Do?

The functions that a NOC currently performs vary widely among IT organizations, but when it comes to how the NOC functions, one of the most suggestive findings is that just under two thirds of the NOC respondents said they believe that the NOC tends to work on a reactive basis, identifying a problem only after it impacts end users.

A CIO who was interviewed stated that the most frequent question he gets from users is, "Why don't you know that my system is down? Why do I have to tell you?" His comment suggests that when end-users notice problems before IT does, it has the effect of eroding the users' confidence in IT in general.

### The most common type of event that causes our NOC personnel to take action is:



## During the past 12 months, our NOC Personnel have spent the...

...addressing issues with...	Greatest Amount of Time	Second Greatest Amount of Time	Greatest Increase in Time
Applications	39.1%	16.9%	45.0%
Servers	14.1%	21.5%	21.7%
LAN	10.9%	15.4%	5.0%
WAN	23.4%	30.8%	11.7%
Security	9.4%	6.2%	10.0%
Storage	3.1%	9.2%	6.7%

### Where Does the NOC Spend Most of Its Time?

While NOC personnel support a broad range of IT functionality, survey results indicated that they are spending the greatest amount of time on applications, which is a relatively new phenomenon. An additional conclusion is that NOC personnel support a broad range of IT functionality.

When analyzing where the NOC spends its time, however, equally interesting is the vast gap in perceptions between those inside and outside the NOC. NOC personnel said that they spend the greatest amount of time on application delivery and performance (39.1%), while non-NOC personnel said that the NOC spends the most time on the WAN (48.3%). This perception gap is supported by the data mentioned earlier indicating that the role of the NOC is not well understood outside of the NOC.

Roughly three-quarters of the respondents indicated that NOC personnel now perform some functions that were previously considered to be Level 2 or Level 3 functions. Another important area of change within the NOC is the shift away from having NOC personnel monitoring management consoles all day, waiting for green lights to turn yellow or red. In particular, over a quarter of the NOC respondents indicated that their company had "eliminated or reduced the size of our NOC because we have automated monitoring, problem detection and notification." It is important to note that in all likelihood, a notably higher

percentage of organizations have implemented automated monitoring but have not eliminated or reduced the size of their NOCs.

### The Next-Generation Operations Center

The survey data collected indicates considerable dissatisfaction with the role currently played by the NOC and also hints at widespread interest in making significant changes to the NOC. It would be easy to create a vision of a next-generation integrated operations center (IOC) that is highly automated, has very effective processes, and is responsible for

“While NOC personnel support a broad range of IT functionality, survey results indicated that they are spending the greatest amount of time on applications, which is a relatively new phenomenon.”



managing the availability, performance, and security of all of the components comprised by IT. Such an IOC would not have to be housed in a single facility, nor would it necessarily have to be provided by a single organization. It is possible, at least in theory, for an IT organization to implement operations centers that have common tools, goals, language, and effective processes independent of geographical or organizational boundaries.

Over the next few years, a small percentage of IT organizations will be able to implement an effective IOC. However, due to a variety of factors, implementing an effective IOC is out of reach for most IT organizations. The lack of effective processes is one of the primary factors currently limiting an IT organization's ability to implement an IOC. By the same token, most NOCs today are somewhat narrowly focused, yet they still suffer from ineffective processes.

It is difficult to believe that over the next few years IT organizations will be able to implement effective processes that transcend a wide range of organizational and technological boundaries, but in spite of myriad inhibitors, change within the NOC is happening and will continue to happen. Below is an inventory of some of the areas where NOCs must focus on in order to evolve to an IOC. We conclude with a prognosis of how successful NOCs are likely to be in each area over the next 18 months.

“Many NOCs have begun the shift away from having NOC personnel sitting at screens all day waiting for green lights to turn yellow or red.”

**Process Improvement:**

There is clear recognition on the part of the survey base that the NOC needs to improve its processes. There is also clear acknowledgement that the vast majority of IT organizations will use ITIL as part of their process-improvement efforts.

**Focus on Performance:**

This focus on performance, as opposed to a focus on availability, is likely to increase, in part because placing greater emphasis

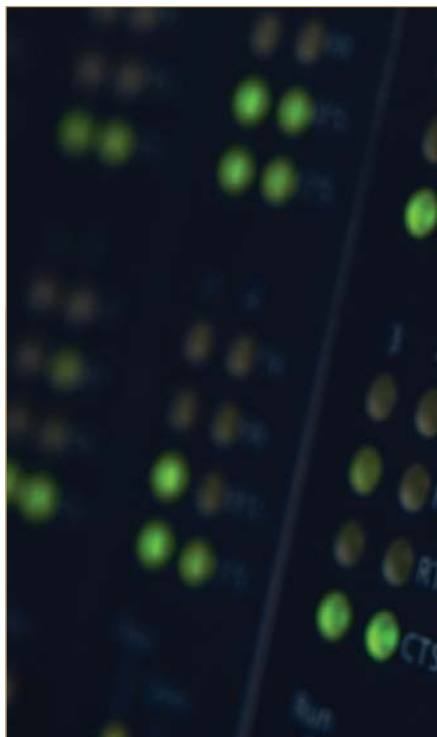
on ensuring acceptable application performance for key applications is the strongest factor driving change in the NOC. However, as strong as the movement is to focus on performance, it is not universal.

**Skilled Staff:**

In general, the skill set of NOC personnel has been increasing, and the majority of NOC personnel are now performing functions that until recently were considered to be Level 2 or Level 3 functions. However, while the skill of NOC personnel has generally been increasing, that is not the case in many organizations.

**Intelligent Tools:**

Many NOCs have begun the shift away from having NOC personnel sitting at screens all day waiting for green lights to turn yellow or red. In addition, over a quarter of the NOC respondents indicated that their company has "eliminated or reduced the size of our NOC because we have automated monitoring, problem detection and notification." This trend, combined with the trend to increase the skill set of NOC personnel, indicates that more intelligence is being placed in the NOC, and that intelligence amounts to a combination of people and tools.



### Tool Integration:

Having a collection of management tools that are not well integrated with each other is currently "a fact of life." Tool integration is one of the biggest issues organizations hope to address with their NOC redesign projects.

### Focus on Applications:

NOCs currently have a significant focus on managing application performance. The survey results also indicated a very strong interest in improving NOCs' ability to manage application performance. However, managing application performance is not always the responsibility of the NOC.

### Security:

Two-thirds of the survey respondents indicated that a growing emphasis on security will impact their NOC over the next 12 months, despite the fact that right now, NOC personnel do not spend much of their time on security.

### Being Proactive:

Today's reality is that the majority of NOCs tend to work on a reactive basis, identifying a problem only after it affects end users. This has been a key issue for a long time and will continue to be a key issue, with only modest improvements being made.

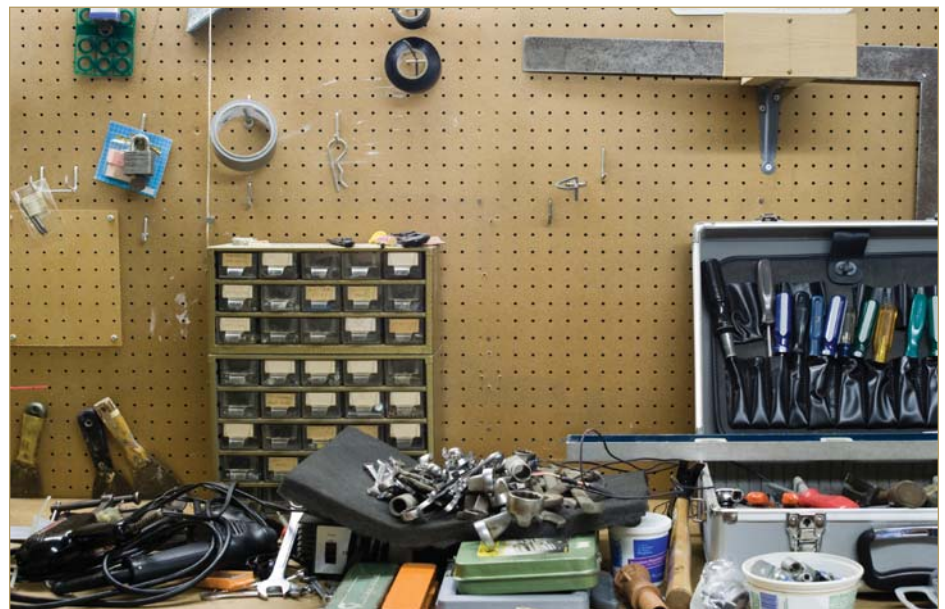
### Conclusions:

As noted, the migration away from today's stove-piped, reactionary NOC to an effective IOC will not be easy. The preceding list identified a number of characteristics associated with an IOC and the likely improvements that IT organizations would make in those areas over the next 18 months.

However, what is still missing in many cases is the vision of senior IT management about the future role of the operations center. In addition, the survey results showed a widespread lack of knowledge inside the IT organization

about the role of the NOC, as well as an accompanying lack of interest on the part of the interviewees to better market the role and contributions of the NOC.

Without this vision and an effective plan to increase awareness and understanding of the NOC and its integral role in the wider enterprise, IT organizations will not be able to make the fundamental changes to the NOC that are required for its transformation.



**HAVING A COLLECTION** of management tools that are not well integrated with each other is currently "a fact of life."

# A CLOSER LOOK

Patrick Ancipink, Director, Product Marketing, NetQoS

## RETHINKING "SAFE" CHOICES:

### ARE INCUMBENT SYSTEMS REALLY WORTH THE COST?

#### Old Technology and the Danger of Incumbency

Ten questions you can use to determine if it's time to replace an IT management product



With rapid adoption of new technology and constant change in IT, we often lose perspective on some of the products we purchased years ago. We may not really understand their true cost, or give much thought to better alternatives. This is particularly true in the case of some older network management products that have become commoditized over the years and hover under the radar of scrutiny.

Yet we all recognize from our own experience that IT professionals are regularly hindered in their daily tasks by legacy management tools that are outmoded, that are increasingly and annoyingly cumbersome to maintain and administer, or that no longer deliver the value for which they were originally acquired.

Each year, companies renew exorbitant maintenance contracts for dormant products from vendors that seem hard-pressed to retain a single trained support representative to answer the phone. Such products have often been acquired as a cash cow, already feature-rich and beloved by their current customers. The new parent vendors immediately begin to starve their development, choosing high margins over more customer-centric benefits like taking advantage of new standards and protocols, or integration

with new architectures or solutions. Other times, these products have just been allowed to wither within a bigger portfolio because the talent and resources of the vendor have been reallocated for more pressing needs. Such is the power of incumbency.

Beyond maintenance and license costs, we also have to consider the opportunity costs related to speed, quality, scalability, and efficiency. The limitations of old technology can not only make the work more tedious, but can also force you to work with incomplete or fragmented data that's hard or even impossible to correlate or integrate with other data—all of which lead to poor decision-making.

For example, many legacy products were designed for an environment in which the key consideration was uptime. However, equipment in use today is generally reliable, and fault is no longer the priority. Products that only tell you whether the device is up or down cannot be the center of a performance or service management approach. They are fundamentally reactive and don't provide the data or analysis required to help you understand declining performance before network users start to complain. When you combine high

maintenance costs with such opportunity costs, the financial penalty for the status quo begins to add up.

No one should advocate tossing out a product or technology just because it's a little long in the tooth—there are some great workhorse products that do more than earn their keep—but we should take a little more time to assess the situation and ask some important questions:

**1** **When was the last time the incumbent product provided an update I cared about? Is there a compelling roadmap for future development?**

If your IT management tools vendor hasn't announced a significant new release within the last two years, it may be time to ask about their mid- and long-term plans for the product. You may uncover some unpleasant truths, such as a sunset date.

“Products that only tell you whether the device is up or down cannot be the center of a performance or service management approach.”

### A Replacement Example:

The Network Engineering staff at a Fortune 100 financial firm in California was fed up with a dated SNMP device monitoring product. As data center consolidation increased the number of applications running over the WAN, network performance was becoming more critical, and the legacy polling product was not providing the early warnings or detailed information the company needed. It wasn't just that the tool was old; any newer functionality was added through acquisition and required different installation and configuration. Integration and workflow with other tools was non-existent, even with the products from the same company. The vendor minimized investments in the product line for a decade, and the users described the tool set as "Frankenstein," with bolted-on additions and halting performance. After countless acquisitions and mergers, the knowledge base of the vendor on this product was nearly useless. First-line support had been outsourced to a very low-value, generalist group that had trouble even capturing the right information that they would then forward to the few people who actually knew something about the product.

When a maintenance renewal proposal arrived from the product vendor with a yearly six-figure charge, the Network team finally said "Enough." They began to look for a product that was more cost-effective and that fit in with their long-term network performance management plans.

The team was delighted to find a vendor offering a replacement product that provided all of the functionality of the old tool, and that added support for advanced performance capabilities (like IP SLA, CBQoS and NBAR) that would help them monitor QoS policies and prioritization. The new vendor also supported the product with an expert, single-tier customer care organization.

The new device management product was part of a comprehensive, actively developed network performance management suite focused not only on the individual components, but also on the operational workflow and integration that the team needs every day. And ultimately, it cost less than a year's worth of maintenance from the incumbent vendor.

**2** Does the incumbent product use older techniques or technologies that overtax the infrastructure or available resources?

Consider whether you are continuing antiquated practices or retaining marginally useful resources—including human resources—just to support older tools.

**3** Am I training new talent on potentially irrelevant tools?

Top IT talent is retained and motivated by opportunities to work with the latest techniques and technology. Smart people like to learn new things. Are your best people sticking around, or do they tend to jump ship when a new challenge presents itself? Hiring and training new people can be very expensive.

**4** Does the incumbent product exploit or support new capabilities in complementary technologies?

Adopt a best practice from Applications teams and think about long-term architecture needs and standardization. For example, will the incumbent tool help you manage an eventual transition to IPv6?

**5** **Is the incumbent product's parent company just milking the product for maintenance revenue, or are they continuing to invest in its development and support?**

Compare what you are getting in return for your maintenance and support dollars across your family of vendors. Inquire about the post-sales company support in categories such as the location and number of support tiers, the skill level of support personnel, problem resolution response time and escalation procedures, and single point of contact account management.

**6** **Could I redirect the maintenance cost of the incumbent product toward acquiring a newer, more relevant product that is actively being developed and improved?**

It can be a challenge to find the time to do the necessary research into competing products, but such research is important. Even if you don't plan to swap out the legacy tool, you need to keep up with the latest advances in the product categories relevant to the tools on which you rely. While you are browsing the most recent industry publications or attending a trade show, collect some data on pricing so you can make a fully informed decision to either keep or deep-six your current toolset.

**7** **Are there any "trade-in" programs available to help with cost and transition?**

Vendors know that switching products is not a decision for the faint-hearted, so they will frequently offer migration programs, with significant pricing advantages, and services tailored for replacement.

**8** **Is there an outsourcing risk because I employ outdated technologies and techniques?**

While outsourcing certainly makes sense in some situations, you want to avoid "punitive outsourcing" that results from neglecting a core competency that could be performed more effectively.

**9** **If challenged by my management, could I justify the cost of the incumbent product against alternatives?**

Take a page from ITIL v3 and Business Service Management; understand your value to the business, in business terms, and be able to back it up with specifics.

**10** **What are the product roadmaps, investment plans, and possible pitfalls associated with alternative products or services?**

Be sure to consider the costs associated with deploying and learning a new tool. And try to get a specific commitment from the vendor for future enhancements to the competing product. These discussions will give you a sense for how deeply committed a vendor is to each product and to its customers.

This last question is perhaps the most critical one to answer. After all, we don't want to swap one diminishing product for another and find ourselves in the same situation all over again.

Often, the decision to replace incumbent technology is postponed indefinitely as there are always new fires to fight and more tantalizing new projects on the horizon. But replacing legacy tools with newer technology that allows you to work more effectively can mean reduced costs for the business. Considering the potential advantages of upgrading hardware, such as productivity increases, quality of work improvement, and importantly, mitigating risk, we should reserve some time to evaluate incumbent products and ensure their vendors are earning the share of IT budget they are taking.

## EBOOK

Jeffrey Hicks, VoIP Architect, NetQoS

## VOIP: DO YOU SEE WHAT I'M SAYING?

Note: The following piece is an excerpt from Chapter 4 of the NetQoS eBook entitled *VoIP: Do You See What I'm Saying? Managing VoIP Quality of Experience on Your Network*, by author Jeff Hicks, NetQoS Software Architect and VoIP expert. This excerpt advises on the strategies to successfully manage unified communications in your IT infrastructure. To read the full chapter, as well the entire eBook, please visit:

<http://www.netqos.com/ebook>.

### Unified Communications

There have been many discussions in the industry centered around how you can manage the VoIP quality of experience on your network. And numerous strategies have been developed to help you estimate the levels of quality that users experience in their interactions with the VoIP phone system and continually deliver a high level of quality. But VoIP is just the beginning. It provides a great starting point on the path to Unified Communications. The network requirements for VoIP, call setup performance, and call quality management are important foundational concepts to build on as you expand the capabilities of your network. Deploy VoIP and get it right; then, you are ready to move on to other Unified Communications applications like video, presence, and unified messaging.

Many enterprises are beginning to take a close look at the substance behind the hype surrounding Unified Communications, or "UC." A recent survey showed that nearly 30% of enterprises had a Unified Communications strategy, and more than 31% viewed Unified Communications as one of their top three IT initiatives. [1] Software heavyweights such as Microsoft® and IBM® have entered the UC market, which was previously owned by the IP PBX vendors, such as Cisco® and Avaya®. As the solutions evolve over the next several years, expect business applications and integrated soft phones to play a greater role and to find much broader acceptance.

As these new communications applications are added to the network, what are the key network performance considerations, and how can you manage them?

UC adoption is usually a slow, staged process and not a "forklift" upgrade. When examining the path toward Unified Communications consider the following:

- What is Unified Communications?
- What are some of the new Unified Communications applications?
- How will a Unified Communications deployment affect network performance?
- How can you manage Unified Communications applications?

Unified Communications offers the vision of great productivity gains as integrated, multi-modal communications interfaces are built into our most commonly used business applications. But anytime new applications are added to the network, it's always good to take a step back and analyze the potential impact on network performance.

“ The network requirements for VoIP, call setup performance, and call quality management are important foundational concepts to build on as you expand the capabilities of your network. Deploy VoIP and get it right; then, you are ready to move on to other Unified Communications applications like video, presence, and unified messaging. ”

Before we do that, let's take a look at exactly what we mean by the catch-all term Unified Communications.

## Defining Unified Communications

Ask the question "What is Unified Communications?"; and you are likely to get many different answers. After sorting through all the marketing messages from the major vendors, how should we define Unified Communications? Let's start by pointing out what UC is not. Unified Communications is NOT:

- **VoIP only.** VoIP-based call processing is a building block for UC, but VoIP alone is not enough to provide UC.
- **Unified Messaging.** The idea of getting all your messages - email, voicemail, fax - in a single interface has been around for some time now. While UM simplifies message access and is generally part of a UC strategy, it is not, by itself, UC.
- **Closed, proprietary systems.** UC depends on interoperability between applications and infrastructure. If you can't communicate with a colleague because she is using a different vendor's communication system, there's not much point in being "unified" in other areas.
- **Rip and replace.** You likely have a communications infrastructure in place already. UC should work side-by-side with your existing infrastructure to enable new applications, not force the replacement of existing infrastructure.

## Now let's define UC from a performance-first perspective:

// Unified Communications is the integration of multiple modes of communication within applications and infrastructure that allows people, teams, and organizations to communicate more effectively. The IP network provides the unifying factor for UC, and network performance is a critical enabler. //

- **About big cost savings.** UC may not save you money. It requires deployment and management of new components and applications. UC vendors often tout user productivity benefits as a cost justification. While we think these benefits can be substantial, it's difficult to put a hard dollar value on soft benefits like these. (Of course, you can always hire an expensive consultant to analyze productivity gains from a UC deployment, but that supports our point about the cost).

There's one common factor when discussing Unified Communications applications: they all make use of a converged IP network. In order to provide benefits from the real-time presence status, point-to-click calling, video conferencing on demand, and other features built into UC applications, the network must be managed and tuned for optimal performance. From the perspective of user experience, UC applications will place greater demands on

your network than any other networked applications to date.

Unified Communications solutions provide applications that allow for communications in a variety of different modes. Let's discuss some of these applications and their impact on network performance.

## Unified Communications Applications

Unified Communications applications are designed to streamline business processes. Communications are a key part of any business, and ineffective or unavailable communications media can directly affect the bottom line. Think about the flow of information through your company. Your business has processes in place to route information to appropriate parties who act on the information and often, in turn, create additional information that must be acted on. These processes are prone to inefficiencies.

## The Business Problem

UC applications strive to improve end-user productivity by addressing the business problem of communications inefficiency. Communications inefficiencies are created in a number of ways.

**Phone Tag.** We've all participated in this little game. You don't know that someone is unavailable or in a meeting, so you call and leave a voicemail message. Then that person calls you back, but doesn't realize that you're now out of the office - so they leave another voicemail. Phone tag results in wasted time.

**Number Lookup.** You get an urgent email from a colleague, and you need to quickly give him a call. Unless you have freakish memory recall, it's a good bet that you'll have to look up the colleague's phone number. Hopefully, your contact information is up-to-date and accessible.

**Switching Applications.** You're working with a business application and need to communicate with a colleague about a report you are viewing. Switching out of the business application and launching an email program takes time and loses the context of your working environment. If you can instead communicate from within the application, you save time and maintain the context for the communication.

**Human Latency.** We all know the effect that network latency can have on application performance. You are working

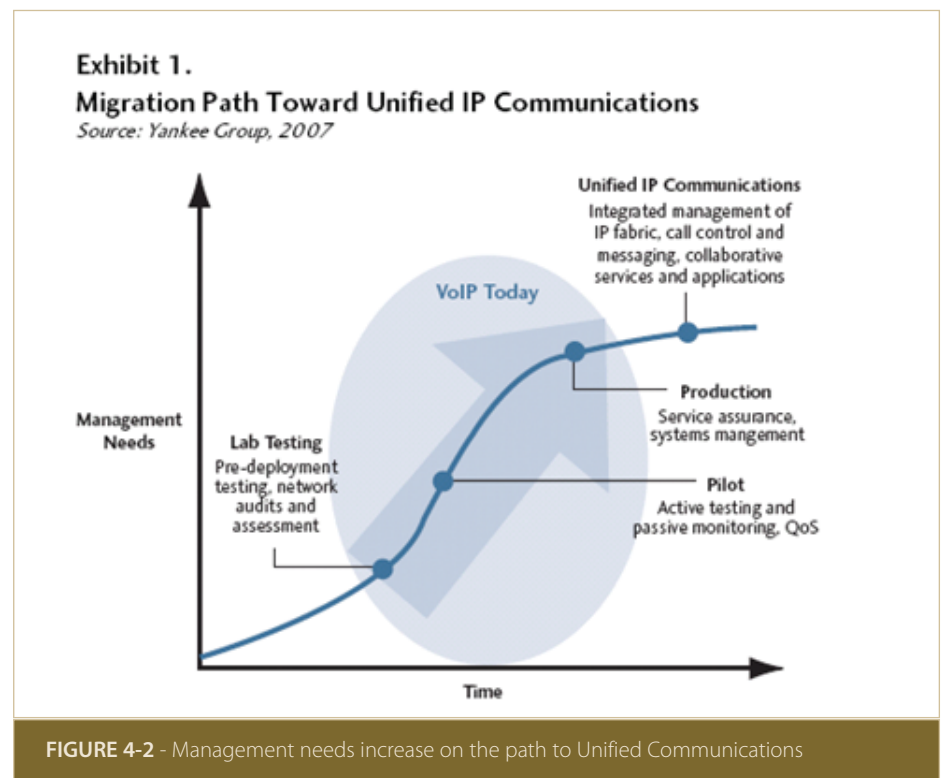
on a project and need immediate input from a supervisor to move to the next step. Unfortunately, she is out of the office, meeting with a client. The time it takes for you to communicate with that supervisor and get a response could be considered human latency.

**Globalization.** The workplace is growing more and more global in nature each and every day. Working on a team that is geographically distributed - with team members in multiple global locations - requires excellent communications.

**UC applications** are geared toward addressing these efficiency issues and providing more effective communications for the enterprise.

## The Network Foundation

A good network foundation is required to prepare for current and future UC deployments. Enterprises will follow different timelines as they begin this journey. But no matter how slowly or how rapidly the deployment proceeds, the need for a sound network management strategy increases as Unified Communications applications are rolled out beyond the initial pilot phase. The 2007 Yankee Group Research report titled "A Guide to Managing Enterprise Unified Communications" offers a look at the migration path from VoIP to Unified Communications and various management considerations along the way. [2]





It's interesting to note that network management plays a key role in every stage of the overall UC deployment cycle. It begins with Lab Testing and continues through Pilot and Production phases - all in preparation for deployment of Unified Communications applications. At each step along the way, it's also important to look at key metrics for your existing applications to ensure that the changes haven't harmed their performance.

Network performance is always an important component of end-user satisfaction with application delivery, but as we've pointed out in previous sections discussing VoIP performance, the network directly impacts the user experience with UC applications. According to Zeus Kerravala, Senior Vice President of Enterprise and Enabling Technologies at Yankee Group Research, "When UC applications encounter performance issues, they don't just slow down, they become unusable. The complexity of presence, voice, and video on your IP network demands that network management tools present a unified view of the performance of these applications."

With that advice in mind, let's take a closer look at the impact that UC applications can have on network performance.

### Unified Communications' Impact on Network Performance

Any UC solution comprises lots of moving parts. Plenty of new applications, new endpoints, and new infrastructure must be

added to your network. As Unified Communications applications are purchased, installed, and configured, prepare yourself for the inevitable changes by finding preliminary answers to several important questions:

- How will the new UC applications affect the performance of my existing networked applications?
- How will the new UC applications themselves perform?
- If performance is sub-par, will the user experience be good enough to make deployment worth the effort and expense?

To answer these questions, it's worth considering the kinds of requirements that typical UC applications can place on the network in terms of bandwidth, packet loss, jitter, and latency. Understanding how the new applications work is important as well. The Session Initiation Protocol (SIP) is the key network protocol that enables communication for almost all UC applications.

SIP is not just a voice call setup protocol. It was designed to be more than that; in many current solutions, it enables real-time communications for audio, video, IM, and presence. Table 4-1 shows how SIP is used for the main applications typically present in a UC solution.

Because SIP is practically ubiquitous when it comes to UC applications, it is important to understand its performance characteristics. For example, one of the advantages of SIP is that it's a text-based protocol, easy to parse and read. But this advantage can also cause network performance issues: being ASCII-based means that many required SIP messages are quite verbose, consuming more bandwidth. We've observed from packet analysis of a standard phone call that the call signaling information from a SIP phone can consume up to four times more bandwidth than a phone using Cisco-proprietary SCCP. Call signaling is not a large bandwidth consumer in general, but when you deploy thousands of endpoints, the extra bytes can add up.

UC Application	SIP Usage
Voice	Set up and take down voice calls
Video	Set up and take down video sessions
Instant Messaging	Establish IM session between users and send/receive user text messages.
Presence	Used to allow endpoints to subscribe to presence status and receive presence change notifications.

**Table 4-1 - SIP is a key enabler protocol for UC applications.**

Another set of performance issues is inherent in the networking architectures that carry SIP data. As more and more UC applications that use SIP are deployed in the enterprise, organizations will need to communicate with other organizations outside their domain. In the past, the PSTN network functioned like glue to tie everything together. Islands of VoIP existed within enterprises, but PSTN connectivity was still required to call other users outside the enterprise. More recently, service providers have begun offering SIP trunks that can connect the SIP islands together and allow SIP-based communications between different domains. A SIP trunk is basically a network connection to transfer SIP packets and data traffic up to an allocated amount of bandwidth.

Deploying UC applications can raise a whole host of new network performance issues above and beyond those associated with the SIP protocol and system architecture. One feature that makes UC attractive is the real-time nature of the communications - it fits the model of how humans like to communicate. But in order for Unified Communications to fit the "real-time" model, you need a network that's ready to provide optimal real-time performance.

## Managing Unified Communications

Managing the performance of Unified Communications applications presents a challenge to the traditional organizational

hierarchies at many enterprises, where separate teams of trained staff have responsibility for different communications components. In the past, a typical enterprise had a group to manage their PBX, another group to manage the network, a different group to manage servers, and possibly even another group to manage specialized infrastructure like video conferencing. The transition to VoIP has initiated the convergence of not only the networks, but also the management groups. Many telecom and data management groups are becoming a single entity. Now with the addition of UC, we see the further convergence and blurring of traditional boundaries.

Successful delivery of UC applications will require a performance-first - proactive quantifying of network and application performance - mentality, applied to the management of the applications and infrastructure. It's not enough to know whether the server is running, or the router is up. UC will cross all the boundaries of application, voice, server, and network management and demands a unified approach to management.

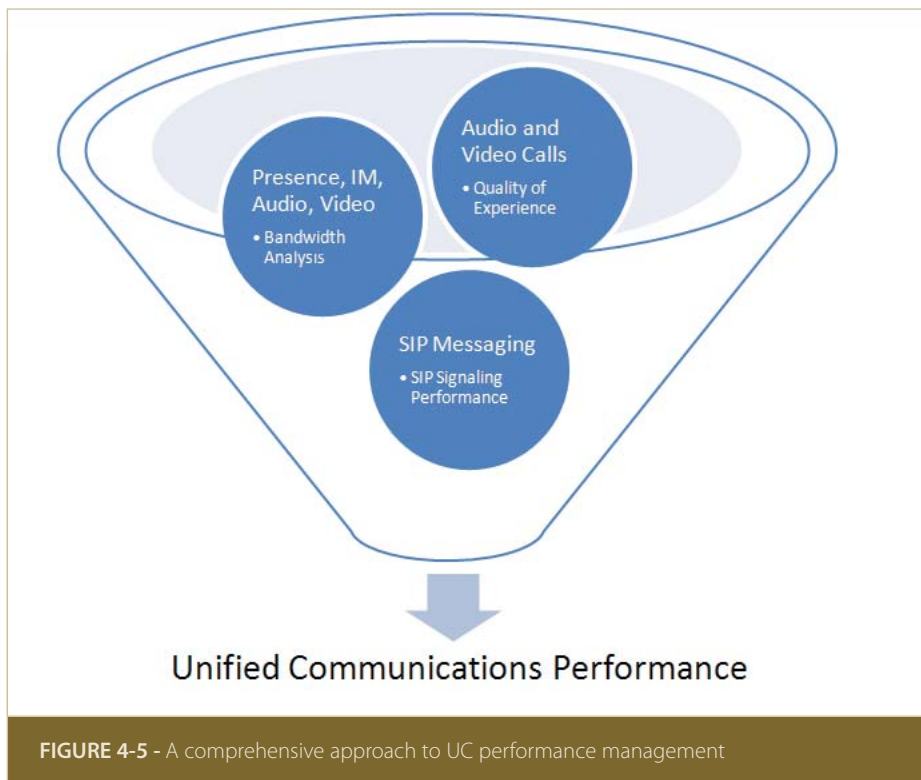
Managing Unified Communications starts with understanding the components that create a better QoE for your users. You need tools to provide the visibility into the QoE metrics and the ability to map those metrics to underlying network quality of service. Perhaps most importantly, you have to do away with the finger-pointing traditionally associated with the separate IT and Telecom teams of the past. Keeping

users happy as you roll out not only new applications, but new ways of communicating with coworkers, both distant and close by, requires a cooperative and comprehensive mindset. Let's consider a unified approach to UC management that prioritizes performance.

## UC Management Methodology

Since UC applications span a broad spectrum in terms of network performance, to manage them all, it's useful to look at a number of different performance data sources and pull them into a single UC Performance "Dashboard"-type report. A performance-first UC management strategy should include performance characteristics for all of the UC applications, always bearing in mind that they have the potential to drag each other down. Each application will have one or more performance metrics associated with it that is best suited for the approach outlined here. As the performance information from the data-gathering tools is synthesized in a single report, it should present a comprehensive management view of the UC environment. Figure 4-5 demonstrates some of the items that should be part of UC performance management.

UC performance management begins with the Quality of Experience. What is the QoE for the user audio and video calls? What is the underlying QoS that supports that QoE? These are key questions that should be answered by UC performance monitoring tools.



For voice, the QoE is best measured by the call setup and call quality performance. For call setup, look at metrics like delay to dial tone, post-dial delay, and call setup failures. For call quality, begin with MOS, but also understand the relevant network metrics like packet loss, jitter, and latency. For video, quality standards are not as well defined as they are for voice. It's hard to use software to quantify something like "lip-sync delay," for example, even though a viewer most certainly notices the effect when the image and audio are poorly synchronized. In the absence of a de-facto quality standard for video, a more user-friendly approach to video quality monitoring relies heavily on metrics that

are measured by the video components themselves, such as video frame loss and frozen video. You can then correlate spikes in these values with the underlying network metrics for packet loss, jitter, and latency. Keeping all of these metrics is as important to video quality as it is for VoIP call quality.

Moving from user experience to bandwidth analysis of UC applications is a logical progression. We identified some of the performance concerns around bandwidth for UC applications like video and presence. The data collection tools you use to manage UC performance need to provide enough visibility into traffic composition on network links so that you

gain the necessary visibility to understand:

- how much bandwidth is being used by each of the UC applications, and
- who is using that bandwidth.

Desktop video has the potential for consuming large amounts of bandwidth in places you don't expect. If a specific user is making video calls all day long to other users and saturating your WAN link, you need to know about it right away. You also need microscopic visibility into every network link where QoS is being applied. A QoS misconfiguration at any point in a communication flow between two users can make VoIP and video nearly unusable as the tiny VoIP packets get queued behind the huge video packets, or as video traffic is queued with other data traffic.

Finally, we touched on the fact that SIP was the underlying enabler for all UC applications. It only makes sense that we should keep an eye on the performance of SIP signaling and message flows. Within the SIP messaging performance, it's important to understand what kind of Network Round Trip Time (NRTT) is typical between client UC applications and your UC servers. What is the server response time, and is it degrading over time?

A network performance product like NetQoS Performance Center can help answer these questions and report on underlying metrics that affect UC application performance.



**FIGURE 4-6** - A Unified Communications dashboard provides comprehensive view of UC performance

## Summary

Unified Communications is another in a long line of new applications that offer compelling features, all of which are accompanied by network performance ramifications. We always recommend that before you deploy a new networked application, first, understand its network performance requirements, and second, understand the potential impact on your existing applications.

As applications like UC have increased user interaction, understanding the user's quality of experience and being able to map that quality to performance levels for management and troubleshooting are crucial. Use monitoring tools to gain the visibility that you need to provide an excellent unified communications experience for your users.

In Figure 4-6, we've created a Unified Communications Dashboard using the NetQoS Performance Center. The dashboard shows at a quick glance the UC Quality of Experience, a UC Bandwidth Analysis, and a view of SIP Messaging Performance.

The different views on the "dashboard" report provide insight into UC performance. The Call Quality Breakdown view provides a breakdown of audio calls and the user experience based on the MOS for those calls. The Call Performance by Location view shows ratings for audio and video performance metrics for calls in specific network locations. Selecting a "location" allows you to drill in and see the metric details, including key call-quality and call-setup metrics, measured on specific subnets. The Performance by

Application view provides ratings for performance metrics associated with SIP messaging. Selecting the "SIP Messaging" applications allows you to drill in to see the details of salient metrics such as transaction time and network round trip time.

Bandwidth analysis is provided by the Top Enterprise Protocols by Volume and Top Enterprise Hosts by Volume data views. With these views, we can see how much bandwidth is consumed by voice, video, and SIP. In addition, we can see who is consuming the bandwidth so that we can make performance-enhancing adjustments. For example, if video conferencing is using up a lot more bandwidth on two or three key links, we might need to move the conferencing server to another location.

## References

1. Kelly, Brent. "Do You See What I See in UC?" Business Communications Review, November 2007. Posted on: <http://www.nojitter.com/showArticle.jhtml?articleID=205901063>
2. Yankee Group Research, Inc. "A Guide to Managing Enterprise Unified Communications," January 2007.

# REALITY IT: | A LOOK INSIDE OSF HEALTHCARE

OSF HealthCare is a multi-state corporation with integrated affiliates that create a network of healthcare facilities providers, known as OSF Medical Group.

Consisting of seven acute-care facilities, one long-term care facility, and two colleges of nursing, it also has a primary-care physician network of 165 physicians and 48 mid-level providers, known as OSF Medical Group. OSF HealthCare owns OSF Saint Francis, Inc., which comprises healthcare-related businesses, and OSF HealthCare Foundation, a philanthropic arm of OSF HealthCare System and OSF Home Care Services. And last but not least is OSF HealthPlans, Inc., a subsidiary of OSF Saint Francis: a licensed managed-care company in the State of Illinois with about 75,000 members.

OSF wanted to implement the latest technologies, including new healthcare applications such as a cardiac Picture Archival and Communications System (PACS), and leading-edge network technologies, such as MPLS and VoIP. They were preparing for a new data center to come online and were working on a server consolidation project.

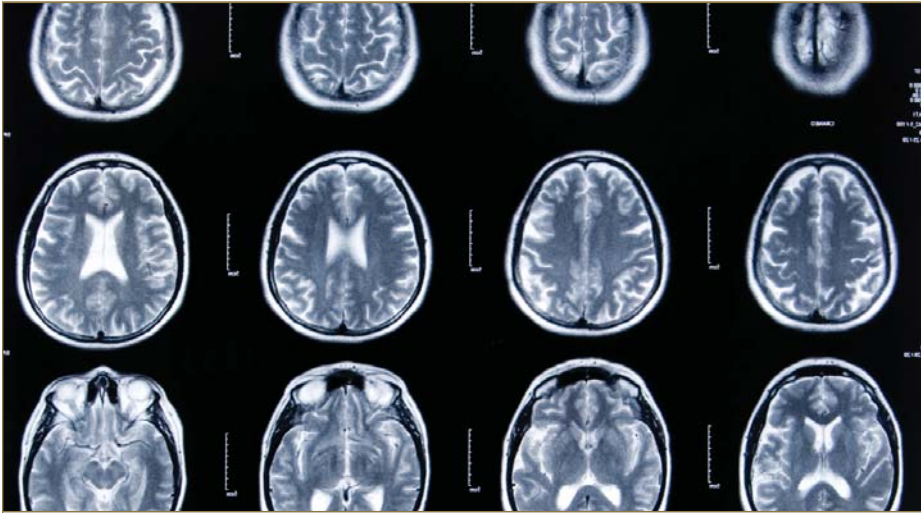
And all of this was the responsibility of a five-member network team, who oversaw network and application services for 10,000 employees in 52 separate facilities.

Maintaining optimal application delivery across the wide area network (WAN) was already a huge challenge for the network team. To prepare for the upcoming changes and address current issues, such as slowdowns with its existing bandwidth-intensive PACS, the network team knew it needed more robust tools to measure and monitor network and application performance.

OSF lacked the organization-wide visibility into application performance, traffic flows, and device performance needed to understand the impact of change, improve end-user services, and optimize the infrastructure. Knowledge of what devices and applications are on the network and how much bandwidth the various applications and users are consuming is critical when rolling out a new application or deploying VoIP. In addition, the network team knew the only way to understand how a change such as a server consolidation impacts performance is to understand how the applications are performing before and after the change.

"When problems occurred, the network was always blamed," said Rob Swain, Network Engineer for OSF. "We had no data to say otherwise or prove what was really going on, such as the ability to pinpoint the cause of an application slowdown. This also impaired our ability to plan for upgrades or new application roll-outs."

The network team chose the NetQoS Performance Center and its three main product modules to provide the network and application performance visibility required to prepare for OSF's new IT endeavors and tackle existing performance challenges. The NetQoS Performance Center is a Web-based management portal that integrates data from NetQoS products in customized views to help organizations be more effective in capacity planning, troubleshooting, and service level management. The three product modules OSF now uses are NetQoS SuperAgent® for end-to-end performance monitoring, NetQoS ReporterAnalyzer™ for Cisco® IOS® NetFlow traffic analysis, and NetQoS NetVoyant® for device performance management.



**EACH OSF HOSPITAL** has a storage area network (SAN) for storing PACS images. Until the network team deployed NetQoS ReporterAnalyzer, they could not determine whether those files were being stored in the right area or how much bandwidth they were consuming in transit.

The NetQoS Performance Center now provides OSF with a top-down view of all applications on the WAN and the ability to drill into the detailed end-to-end performance, traffic analysis, and device performance metrics provided by NetQoS SuperAgent, NetQoS ReporterAnalyzer, and NetQoS NetVoyant. Using the NetQoS Performance Center and its integrated product modules helps OSF deliver consistently high service quality to end users and mitigate the risks of change. They can now measure end-user application response times; understand how changes affect network and application performance; isolate performance problems to the application, server, or network; identify the applications and users consuming bandwidth; avoid unnecessary WAN costs; manage the

convergence of voice, video, and data; and identify virus or denial of service attacks as soon as they occur.

"Without the NetQoS Performance Center and its three underlying products, it would be difficult to get the full picture of what's going on in the network. It is a very powerful tool," said Swain.

### Improved Troubleshooting with the NetQoS Performance Center

The NetQoS Performance Center has helped the OSF network team troubleshoot issues more effectively, especially with its PACS application. Doctors had been complaining that images were taking too long to download, making it harder for them to analyze cases and interact with patients. NetQoS

SuperAgent reports showed that the PACS application was slowing down due to excessive server response times. Each OSF hospital has a storage area network (SAN) for storing PACS images. Until the network team deployed NetQoS ReporterAnalyzer, they could not determine whether those files were being stored in the right area or how much bandwidth they were consuming in transit. Using NetQoS ReporterAnalyzer, the network team found that some images were being sent to different sites across the WAN and not being stored locally, slowing down retrieval times considerably for doctors and unnecessarily consuming large amounts of bandwidth. Once they could see the paths being taken, the network team was able to clean up this process and ensure that all PACS images were stored at the local hospital SAN for faster retrieval. SuperAgent data then verified that the PACS application was performing faster at each site after this fix.

In another instance, NetQoS ReporterAnalyzer alerted the network team that an antivirus program was using large amounts of bandwidth across workstations reserved for PACS traffic, due to a workstation configuration error. Rectifying this problem also freed up bandwidth so that the doctors could access the images more quickly.

"There is so much bandwidth going back and forth between the hospitals and data centers. Now, we have insight into what traffic is traversing the WAN, including conversations and hosts, and how much bandwidth that traffic is consuming. No other product we evaluated came close to providing all the information ReporterAnalyzer does. In addition, with SuperAgent, we are able to see application, network, and server latency as well as set thresholds and baselines," said Swain.

### **Application Roll-outs and Infrastructure Changes: Mitigating the Risk**

As the OSF network team deploys new applications and tackles infrastructure projects, the NetQoS Performance Center provides the metrics needed to mitigate the risks inevitably associated with change. For example, OSF is moving to a cardiac PACS system next year, which is even more bandwidth-intensive than their current PACS system. The network team will use the NetQoS Performance Center to estimate the amount of bandwidth this new system will require and ensure that it is performing optimally without dragging down other critical applications.

In addition, OSF has begun to use the NetQoS Performance Center to plan for its move to MPLS. The NetQoS Performance Center enables the network team to monitor the performance of the infrastructure both before and after

deploying MPLS to gauge the impact of the change. Before the move, OSF is using the NetQoS Performance Center to measure network latency for locations that will be affected; baseline the performance of business-critical applications; configure automatic alerts when performance degrades from normal; identify all applications consuming the bandwidth, who is using them, and when; and monitor router, switch, and interface performance for devices that will be impacted by the move to MPLS.

Once MPLS has been deployed, the NetQoS Performance Center will help OSF ensure that network latency has either improved or remained the same after the change and validate the impact of MPLS on business-critical applications. The team will receive automatic alerts when performance deviates from baselines. With data from NetQoS reports, they'll be able to determine the root cause of any performance degradations, ensure that the QoS policies provided by their carrier are having the desired effect, and monitor router and interface performance for the edge sites after the move to MPLS.

OSF has also started a phased approach to VoIP. Using the NetQoS Performance Center and NetQoS NetVoyant to monitor VoIP performance before, during, and after deployment, OSF will be able to gauge the performance of the network by measuring latency, jitter, packet loss, and MOS via Cisco IP SLA, determine the volume of VoIP traffic across the WAN to make sure VoIP traffic is not starving out other business-critical applications, and measure response times and collect detailed data on any changes to business-critical application behavior due to VoIP bandwidth consumption.

### **Team-Based Reporting with the NetQoS Performance Center**

To monitor overall network and application performance on a daily basis, OSF has taken advantage of the custom reports feature in the NetQoS Performance Center. For example, the Network team has built pages specifically for the Applications teams to understand how their various applications are performing, including PACS and the OSF patient information system and payroll applications.

“Once MPLS has been deployed, the NetQoS Performance Center will help OSF ensure that network latency has either improved or remained the same after the change and validate the impact of MPLS on business-critical applications.”



**THE NETWORK TEAM** has also created reports in the NetQoS Performance Center for the IT teams in each OSF region. OSF has six regions, each made up of at least one hospital and remote physician and business offices (such as the OSF HealthPlans insurance division).

The Network team has also created reports in the NetQoS Performance Center for the IT teams in each OSF region. OSF has six regions, each made up of at least one hospital and remote physician and business offices (such as the OSF HealthPlans insurance division). A report is automatically e-mailed daily to each region, providing an overview of network performance that includes bandwidth utilization, errors on links, and latency of key applications. These reports are also useful for trending, according to Swain.

In addition, the Network team has created its own customized views into network and application metrics. For instance, Swain has a home page he views each day that gives him an overview of router and switch performance.

The Network team's longterm plans include rolling the NetQoS Performance Center out to the OSF Service Center/Help Desk team, with the eventual goal of replacing HP OpenView. Using the NetQoS Performance Center will help the Service Center staff focus on performance instead of just fault management, providing faster

and more efficient troubleshooting by identifying the correct group to call when an issue arises. Swain explains, "The NetQoS Performance Center provides more information than OpenView. Our Service Center people are basically looking at up or down right now. They issue a ping and then call us. When we deploy the NetQoS Performance Center for them, they will be much more proactive in troubleshooting, freeing up our time to work on new deployments and infrastructure upgrades/changes," said Swain.

### **Performance First**

With the right set of tools from NetQoS, the OSF Network team is now taking a performance-first approach to network management. When asked how the use of NetQoS products has improved their own performance, they credit their new-found visibility into the systems they maintain: "Now that we have gone through and discovered what is on the network, we can concentrate on performance, optimizing the network and making it more efficient for delivering application services," says Swain. "Using the NetQoS Performance Center will help us improve and maintain performance as we roll out new applications, move to MPLS, consolidate servers, deploy VoIP, and bring a new data center online," he predicts.



# NetQoS NetAnalyst® Certification

Results-Driven Training for Today's Network Professionals

Monitor, Analyze and Resolve Complex Networking Issues With Network Performance Management Training & Certification



## Topics you'll explore:

- » End-to-end communications within an IP network
- » Protocol dependencies for network and application performance
- » Subnetworks specification and performance management for IPv4 networks
- » Application performance optimization
- » A look at the "carrier cloud" and its effect on MAN/WAN performance
- » MAN/WAN technology selection for remote connections
- » The Planar Troubleshooting Methodology
- » Network instrumentation

## 2008 Course Schedule

These courses fill up quickly, so reserve your seat today:

### NetAnalyst I: Network Performance Technologies

An introduction to network forensics covering core theory and technologies.

October 21-23

Austin, TX

### NetAnalyst II: Network Performance Metrics & Analysis

Move beyond the basics to monitor, analyze, and resolve complex networking issues.

October 28-30

Austin, TX

**Customized onsite training is also available. For more information on course dates visit [www.networktraining.com](http://www.networktraining.com).**

**Standards-based  
Training Drawn  
from Real-world  
Experience**

## 25% Discount on First Course Registration

Register today and mention this ad in the Performance Edge Journal to receive a 25% discount on your first course registration (\$750 value).

[www.networktraining.com](http://www.networktraining.com) | 1.866.764.5278

This offer is exclusive to readers of Performance Edge Journal. Discount applies to the regular course price of \$2995.

# TECH BRIEFS:

Ben Erwin, Technical Marketing Manager, NetQoS

## DETECTING THREATS TO OPTIMAL NETWORK PERFORMANCE

The biggest threat to application performance is change, but today, with the rate of change in the network infrastructure often ranging beyond human scale, it's virtually impossible to track all changes, whether planned or ad hoc, malicious or benign in intent.

**People who manage the delivery of applications need visibility into changes in network behavior to mitigate risks from these changes. Fortunately, modern anomaly detection capabilities can uncover abnormal patterns and pinpoint the source of potential problems before they impact end users.**

In network management terms, anomaly detection is simply determining when network behaviors change from normal patterns. Anomaly detection and mitigation products have traditionally been used by Security Operations and Threat Management teams to detect worms, malware, DDOS attacks, and any unwanted intrusions. The capability has even spawned a new industry term over the past few years: network behavior analysis, or NBA.

To date, the NBA vendors have sold their products to IT security teams with a promise to keep them ahead of the bad guys. This positioning tends to pigeonhole anomaly detection as yet another security solution that is irrelevant to other areas of IT. While there is great

merit in using network behavior analysis for security purposes, this type of visibility can also serve as an early warning system for any network-related performance degradation and should be considered a key requirement for network performance management and application delivery.

Properly designed, anomaly detection products can bridge the gap between Network Operations and Security Operations teams. The products should focus not only on early warnings of security threats, but also on non-malicious application or user behavior that poses a threat to application delivery. In addition, an integrated workflow between anomaly detection and network performance monitoring is essential for network teams. Understanding an anomaly is important, but Network Operations must also be able to assess the impact on application response times and on the network infrastructure. An anomaly detection product that provides detection, impact analysis, and troubleshooting capabilities in a seamless workflow, via a single management platform, should be valuable to any Network Operations group.

To expand their market opportunity, some NBA vendors have added new features to their products and repositioned them for network performance management. However, their solutions generally lack the end-to-end visibility, enterprise scalability, and contextual data necessary to be considered complete network performance management systems. No single metric is adequate for managing application delivery, so it is essential to provide a context around which to integrate and correlate data from multiple sources. For example, if only the host information pertaining to a detected anomaly is presented without data from the associated interfaces and routers, critical dependencies can be missed, masking the impact of the anomaly and making it harder to troubleshoot. Without visibility that extends beyond changes in network traffic, traditional NBA solutions provide only a small piece of a complete application delivery management solution.

Recognizing the need to identify and visualize real-time changes in network behavior, NetQoS has now integrated an anomaly detection capability into the NetQoS Performance Center, a comprehensive, Web-based network monitoring and management console. Unlike stand-alone NBA products, the NetQoS anomaly detection capability is part of a comprehensive management platform that Network Engineering and Operations teams around the world already depend on to sustain and optimize their networks for application performance.

### Anomaly Detection from a Performance Perspective

Several methods and data sources could potentially be useful for detecting anomalies on the network. Most NBA products analyze network flows and packets. Using advanced algorithms, NBA products can study and profile traffic patterns associated with any host (client or server) on the network. Visibility into this traffic can flag a number of different anomalies, ranging from a variation in the

“ NetQoS anomaly detection capability is part of a comprehensive management platform that Network Engineering and Operations teams around the world already depend on to sustain and optimize their networks for application performance. ”

types of packets a specific host is sending over the network to a change in the quantity of packets that host is sending. Either indicator could be an early warning sign of an infected host. But either indicator could equally be a warning of non-malicious user behavior that could nevertheless affect the performance of network applications.

A sudden increase in packet volume could indicate that a user is hosting a non-sanctioned application, such as Bit Torrent or Kazaa for file sharing. This type of change in network behavior is usually not malicious like a worm or virus, but it can threaten application delivery by choking off bandwidth and consuming other resources.

Sudden changes in the types of packets being used by a host may also be an early warning of potentially threatening behavior. Anomalies like these can indicate an improperly configured application that impedes performance if client requests are not being processed properly. Similarly, packet fragmentation is another change in the packet makeup that could be caused by a malfunctioning network device.

Changes to the network configuration can also cause degradation in application performance. For example, if sources of null routes are detected, inconsistent access control lists (ACLs) may be responsible. Monitoring the TTL bit in



**A SUDDEN INCREASE IN PACKET VOLUME** could indicate that a user is hosting a non-sanctioned application, such as Bit Torrent or Kazaa for file sharing.

// When an anomaly is detected, network engineers receiving the alert can access more details in the NetQoS Performance Center and quickly determine, from a single report, how the anomaly is impacting application delivery. //

network traffic can also identify routing loops that are occurring in the network. Detecting both behaviors is crucial to properly securing the network and delivering application services promptly.

More nefarious network activities can create additional anomalies in network traffic composition. For example, fragmented packet sources, SYN-only packet sources, and high packet fan-out may indicate that hackers are attempting to bypass firewalls, or even the presence of viruses and worms on the network. Malicious or not, these types of behaviors can negatively impact the delivery of application services. More packets on the network may ultimately lead to congestion and accompanying packet loss, especially if the non-sanctioned application is heavily used, and, like unexpected packet fragmentation, should be cause for concern to those responsible for managing application delivery – not just to Security staff.

### **NetQoS Anomaly Detection: A Performance-First Approach to Network Behavior Analysis**

By adding anomaly detection capabilities to NetQoS ReporterAnalyzer, the traffic analysis product module of the NetQoS Performance Center, NetQoS provides visibility into changes in network traffic patterns that supports efforts to manage application delivery and secure network resources. Using Cisco IOS® NetFlow as a data source, ReporterAnalyzer provides real-time and historical reporting and analysis for hundreds of thousands of network links around the globe. In addition, its anomaly detection capabilities analyze and profile traffic patterns in the NetFlow data it collects for every client and server on the network. When unusual patterns are detected, network staff are alerted via the NetQoS Performance Center Web portal, by email, or by an SNMP trap.

The NetQoS Performance Center provides a holistic approach to managing application

delivery by correlating an anomaly's impact on application response times, VoIP quality of experience, and device performance. When an anomaly is detected, network engineers receiving the alert can access more details in the NetQoS Performance Center and quickly determine, from a single report, how the anomaly is impacting application delivery. This integrated workflow between behavior analysis and network performance management is unique to the NetQoS Performance Center.

For its core functions, NetQoS Anomaly Detection leverages hallmark features of the NetQoS portfolio: baselines and thresholds. Without a sophisticated understanding of normal traffic patterns on a given network, anomaly detection is tantamount to guesswork, and as a result, it generates too many false positives to be useful. NetQoS has developed a broad set of sensors to build a baseline of "normal" network activity and thus differentiate between common and less common sources of performance-inhibiting network behavior. Using proprietary algorithms to continuously monitor traffic flow patterns for deviations from the baseline or to detect violations of user-defined thresholds, NetQoS Anomaly Detection can simultaneously alert on and report abnormal behavior. In addition, NetQoS Anomaly Detection provides a means of quickly mitigating the risk of anomalous traffic by automatically generating remedial ACL configurations, which can be applied to the hosts generating the undesired activity.

# SECURITY CHECK



**WHILE MANY NETWORK ANOMALIES** are self-inflicted, it is just as important to understand their potential impact on network resources and on application delivery as it is to detect and mitigate external threats to network security.

## Looking Forward

While analyzing network traffic for anomalies is a logical place to start, even more can be done to help network professionals be more proactive in mitigating risks to application delivery. Understanding anomalies in response times and VoIP call quality correlated with network flow patterns can identify the source of anomalies, with insight into the resources that are impacted, and why.

Detecting anomalies in device performance provides an even more granular picture of how network services have been affected. With proven data collection, analysis, and reporting capabilities for response times, VoIP quality of experience, network traffic flows, and

device performance, NetQoS is in an ideal position to broaden the scope of anomaly detection beyond traffic analysis. The integration of anomaly detection into the NetQoS Performance Center provides unique capabilities that arm network operations and engineering staff with the means to detect anomalous behavior and take remedial action before users are impacted.

## Summary

While many network anomalies are self-inflicted, it is just as important to understand their potential impact on network resources and on application delivery as it is to detect and mitigate external threats to network security. Using anomaly detection to detect change in real time is an effective way to mitigate the risk from unexpected activity and ensure end user productivity. NBA products highlight anomalies in network traffic to aid security efforts. NetQoS takes a more comprehensive approach, integrating anomaly detection in its network performance management suite. The NetQoS approach to anomaly detection gives network teams a holistic view of network anomalies that impact application delivery throughout the enterprise.

// With proven data collection, analysis, and reporting capabilities for response times, VoIP quality of experience, network traffic flows, and device performance, NetQoS is in an ideal position to broaden the scope of anomaly detection beyond traffic analysis. //

## ARTICLE

Brian Boyko

## TECHNOLOGY ISSUES IN THIS ELECTION YEAR

Politics and network engineering don't usually mix, but there's no doubt that U.S. technology policy affects U.S. technology companies.

As the 2008 U.S. Presidential Campaign is in full swing, there are a number of issues that should be prominent in the mind of IT workers when they go to the voting booth this fall. For example, network neutrality and broadband policy will affect those companies hoping to roll out SaaS solutions; H1B Visa policies will affect the tech job market and the pace of innovation; and of course, more fundamental questions about data security and privacy have become political issues over the past decade.

Different technical publications have endorsed - or hinted that they would endorse - specific candidates for elected office this year. Whether or not these endorsements will carry any weight is yet to be seen, but that doesn't mean that technology issues aren't worthy of consideration. And if the people who actually know something about technology don't speak up, we could be left with policies shaped by the talking heads on cable news shows who have trouble understanding even basic computer concepts, let alone knotty technology issues, such as network neutrality.

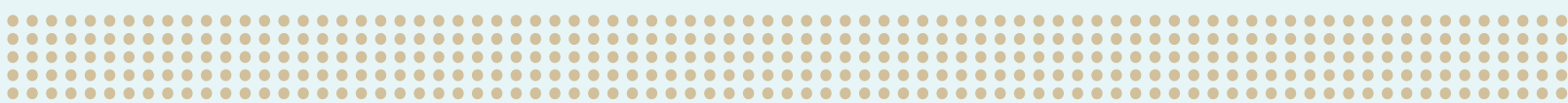
During the main election season, technology issues will probably not be the foremost in voters' minds, but we believe that these elections are extremely important for those who think that a solid technology policy is important to U.S. national prosperity. While we'd feel uncomfortable endorsing any particular candidate, we've put together a list of the top eight current technological controversies that you should consider before voting.

### 1) Intellectual Property Laws

No segment of the technology industry is untouched by the intellectual property laws, both current and proposed. First, any company that makes software, either for resale or in-house use, has to be aware of their rights under copyright law to preserve their own products. Any company that uses - in whole or in part - open-source software needs to be aware of how open-source licenses work; that is, open-source code remains under the copyright of the author, who might be very specific about who may or may not use the license.



**DURING THE MAIN ELECTION SEASON**, technology issues will probably not be the foremost in voters' minds, but we believe that these elections are extremely important for those who think that a solid technology policy is important to U.S. national prosperity.



In addition, the current entertainment industry crackdown on pirated materials affects enterprise networking in a number of ways. First, there's the question of the liability of an enterprise when an end-user on the corporate network uses it to distribute material without the permission of the copyright holder. In a few instances, the new laws have been interpreted such that if you aggressively patrol your network for copyright violations, you can find yourself liable if a violation slips through. This legal gray area leaves enterprise networking in a precarious position: either police the network and assume the legal liability, or take sanctuary in "safe-harbor" provisions and allow the traffic of illicitly traded files to clog up your network.

The current legal climate fosters confusion, but there is also a middle ground. Certain types of traffic can be prevented from taking up bandwidth necessary for business applications by methods that use QoS policies without looking at the individual files, and that seems to be the best solution right now. However, the point here is that any changes to copyright law would have a profound effect on the ways that technology companies do business, and that is why everyone in IT should be keeping an eye on this issue.

## 2) Broadband Penetration/Infrastructure

American broadband infrastructure is simply not up to the standards of other countries. Japan, Korea, and France are

“The U.S. population density may be 31 people/km<sup>2</sup> compared to France's 113/km<sup>2</sup> or 337/km<sup>2</sup> for Japan, but while California has a population density of 90.27people/km<sup>2</sup> - rivaling France - it does not have France's broadband speed. Considering that California is one of America's technological "bread baskets," this is a serious problem.”

often touted as having much better broadband than the U.S., with various explanations given regarding a lower population density in the U.S. However, there's little correlation between population density and broadband penetration when you look at individual states.

The U.S. population density may be 31 people/km<sup>2</sup> compared to France's 113/km<sup>2</sup> or 337/km<sup>2</sup> for Japan, but while California has a population density of 90.27people/km<sup>2</sup> - rivaling France - it does not have France's broadband speed. Considering that California is one of America's technological "bread baskets," this is a serious problem. On the other coast, New Jersey has a population density of 438/km<sup>2</sup>, and New Jersey's broadband speed is no better than that of the rest of the nation. In addition, Norway, Sweden, and Finland have lower population densities and both have faster broadband speeds and greater household penetration than the U.S.

Just as the highways developed by the Eisenhower administration helped to foster America's post-war manufacturing boom, better broadband infrastructure can help improve America's technology industry. A ubiquitous, high-quality broadband network can mean that more applications can be run as a Web service out on the Internet instead of over the WAN. More bandwidth for everybody means that the bandwidth for your company becomes cheaper, and that you can afford more of it, which means that existing apps will run faster – assuming there aren't other network performance problems. And with more bandwidth, you'll be able to run high-bandwidth applications such as Cisco Telepresence.

Even if your company is lucky enough to be sitting on a large amount of dark fiber, every company relies on smaller companies as vendors, as suppliers, as distributors, as customers – and those smaller companies are relying more on software as a service (SaaS) solutions. In

“How much data should the government be able to collect?” and “How well can data the government collects be kept secure?”

the grand scheme of the business world ecosystem, communications infrastructure policy can have far-reaching effects.

### 3) Spectrum Regulation/Allocation

When people think of bandwidth, they often think of bits traveling down pipes. But the other type of bandwidth is just as important: the bandwidth of the electromagnetic spectrum. Because you can't run two different signals on the same frequency (they would interfere with each other), the U.S. Federal Communications Commission (FCC) allocates frequencies and decides which frequencies are going to be used for which purpose. Because the spectrum is limited by physical laws, because some frequencies are better suited for different purposes, and because a huge amount of money is involved, the rules governing spectrum allocation have become bones of contention.

For example, in early 2008 there was an ongoing auction for the 700MHZ band - a slice of the electromagnetic spectrum that can penetrate walls and cover a very wide area. These qualities made this portion of the public spectrum very desirable for the television stations that now control the bandwidth, and also very desirable for cell

phone companies currently bidding for the bandwidth that will be available when the television stations must return it to the FCC as part of the analog/digital TV switchover in 2009.

Anything that deals with broadcasting of any sort - wireless networking, WiMAX, even telecommunications ownership - goes through the FCC, making it one of the most important and powerful federal commissions. Decisions made by the FCC can affect any rollouts your company makes regarding wireless networking or cellular technology, not to mention product offerings designed for certain broadcast behavior.

### 4) Network Neutrality

The possibility of network neutrality legislation, or the actions of big-business players in the absence of network neutrality legislation, can mean fundamental changes in the way that bits travel over the wire. We won't delve into the associated issue here, as network neutrality is covered at length in other publications and debated exhaustively on the Internet. However, the uncertainty over network neutrality is the most disconcerting aspect of the whole debate.

While you can plan for a neutral Internet or an Internet subject to reasonable class-of-service distinctions, it is much harder to prepare contingency plans while this matter remains up in the air.

Some candidates for public office have expressed support for network neutrality legislation, others opposition, and still others ambivalence. And depending on which position is the best for you and your company, the issue is something to consider.

### 5) Communication Interception, Security, and Privacy

Whether or not it is justified, we know that it has been the policy of the current government to intercept communications without warrants, and that some major telecommunications players have helped the government to do so. A number of people are very upset about government wiretapping policy and practice, and the U.S. Congress has helped to thoroughly politicize the debate. The recent debates over wiretapping and other types of government snooping revolve around two core questions: "How much data should the government be able to collect?" and "How well can data the government collects be kept secure?"

Regardless of the supposed value in "fighting terrorism," and excluding, for the moment, any "unreasonable search and seizure" or "due process" issues, these warrantless wiretaps create a third party that is privy to any confidential data that



travels along the wire. While some may trust the government to keep data confidential to the best of its ability, that guidance alone may not be very reassuring to you or to the people at your company who deal with things like Non-Disclosure Agreements and company secrets. We won't venture into political hot water here, but we will suggest that plenty of government agencies are using outdated computers and poorly designed networks, and that those who don't have much faith in the government's ability to keep confidential data secure and confidential can probably cite some excellent reasons for their mistrust.

The U.S. government already collects an alarming amount of data through more traditional methods - from social security records, tax records, and the like - and this information also needs to remain confidential. Computer security policies that are effective, enforced, and adhered to are crucial.

## 6) Open Government Initiatives

One of the ways to increase transparency in government is to make information that the government collects available to the public in an easily computer-parseable, standard format. Some candidates for public office in 2008 have made this a priority, while others have ignored it. The idea is that if government data is online and both easily searched and easily cross-referenced, citizens can use that information effectively.

There are, however, privacy concerns that accompany an open government. In addition, any move to standards begs the question of "which standards?" A related question is whether your IT department will need to conform to those standards in order to interoperate with governmental computers.

## 7) Energy Policy

While energy policy will be a key part of the campaign this year, most people won't be thinking about the effect of

government energy policy on IT departments. But if you manage the budget for the IT department at any substantial enterprise, you know that energy costs directly affect the operating budget. Energy is required to power and cool racks and racks of servers, switches, and routers - not to mention the end-user PCs distributed throughout the enterprise. An election year is prime time for new, sweeping proposals that feed into public concerns, such as high energy and gasoline prices. Various candidates have proposed a carbon tax, for example, which could help reduce atmospheric CO2 but raise energy costs. Other candidates propose government incentives for developing alternative fuels and reducing U.S. dependence on oil, but the global system of energy production and distribution is so complex that even if such initiatives became law, energy costs could increase naturally, for other reasons. Higher energy costs for the IT Department could mean reduced spending in other areas, such as staff, hardware, and software.

By encouraging and perhaps subsidizing the development of computer processors that consume less energy, government can have an additional impact on IT departments. A more subtle effect can be found from the military-industrial complex. Technologies that get their start in the U.S. military often eventually find their ways into the private sector. A military demand for low-power-consuming technology, such as chips and routers to be used in small, autonomous devices, may eventually result in low-power-consuming data centers.



**THE CAMPAIGN THIS YEAR**, most people won't be thinking about the effect of government energy policy on IT departments.

“Obviously, candidates for public office stand to benefit from a crash course in political issues that affect the technology sector of the U.S. economy.”



### 8) Immigration and Education

Immigration is a double-edged sword when it comes to IT, and nowhere is this clearer than with the U.S. H1B Visa program. Supporters of the program claim that the H1B Visa allows the best and brightest of the world's technological geniuses to work for and strengthen companies within the United States. But the number of visas required by U.S. companies in a given year is always up for debate. Too few, and companies are starved for brain power; too many, and domestic IT workers begin to feel the pressure of competing against immigrant labor. The H1B quota has a large impact on both the IT job market and the U.S. technology industry as a whole.

One of the concerns associated with the demand for H1B Visa-holding immigrants is the idea that we do not train enough competent technology professionals in the United States to fill the demand of large companies since computer science and engineering went from the "guaranteed good career" major in the late 1990s to a career to be avoided in the Post-Dotcom Era. Through grants, scholarships, and loans, the federal government has a great deal of influence on what people choose to study.

There have been some interesting proposals to remedy the personnel shortfall, including training those with Associate's Degrees from junior colleges to prepare IT "operators" who lack a full four-year degree in computer science. As hiring and retention are always important concerns for CIOs, a possible labor shortage and policies intended to address it are significant election-year considerations.

### Conclusions:

Obviously, candidates for public office stand to benefit from a crash course in political issues that affect the technology sector of the U.S. economy. Each locality has its own set of policies to debate, as when candidates for state governor are raked over the coals for offering huge tax incentives to high-tech companies who agree to create a few good jobs in their state. Many such issues are being debated, and many more need to reach the political limelight during this election year. We've provided an overview while surely missing several issues. But we hope that this list will help you sort out some of the more significant issues affecting information technology and make better-informed choices when you head into the voting booth this fall.



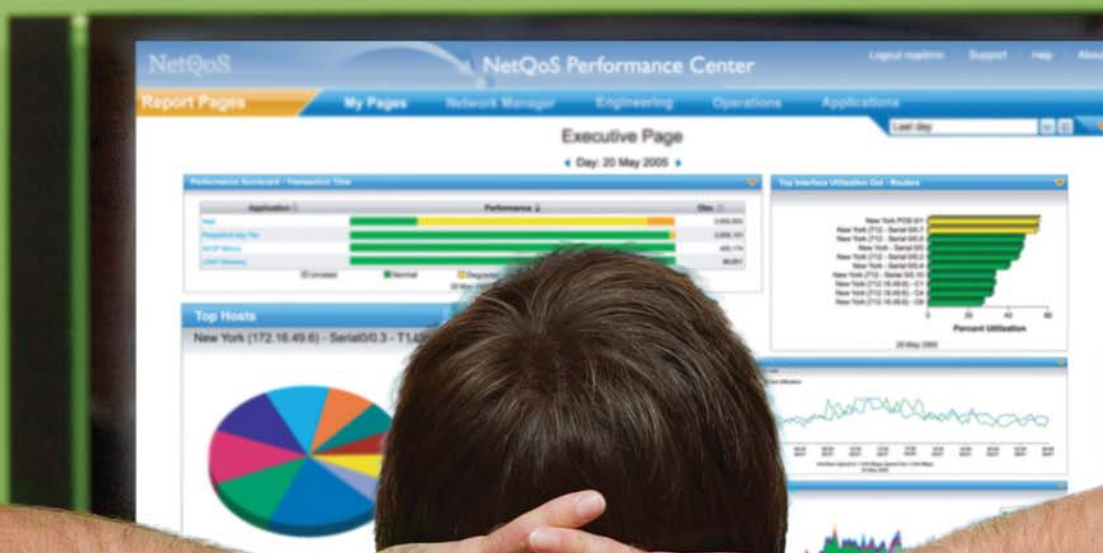
# See what you've been missing with the **NetQoS Performance Center**

## **Gain complete visibility into your network.**

- ✓ Measure end-user to-end application response times
- ✓ Provide consistent application service delivery
- ✓ Correlate network performance to VoIP quality metrics to ensure VoIP quality of experience
- ✓ Understand how infrastructure changes affect network and application performance
- ✓ Isolate performance problems to the application, server, or network
- ✓ Report on network traffic using NetFlow data
- ✓ Avoid unnecessary WAN costs
- ✓ Manage the convergence of voice, video, and data

Sign up for a free demonstration of NetQoS Performance Center at  
[www.netqos.com/solutions/NPC](http://www.netqos.com/solutions/NPC).

[www.netqos.com](http://www.netqos.com) | 877.835.9575 US | 44 (0) 118 929 8032 UK



# BEST PRACTICES:

Dr. Jim Metzler

## 2008 APPLICATION DELIVERY HANDBOOK EXCERPT

### EXCERPT: THE 2008 HANDBOOK OF APPLICATION DELIVERY: A GUIDE TO DECISION MAKING

The IT organization is transforming itself from a loosely connected set of isolated functions - devices, networks, servers, storage, databases, security, operating systems - to an environment based on the recognition that IT is comprised of just two functions, application development and application delivery, and that these functions must work in an integrated fashion for the IT organization to ensure acceptable application performance. This view of IT affects everything - including the organizational structure, the management metrics, the requisite processes, technologies and tools.

**We've provided an excerpt from "The 2008 Handbook of Application Delivery: A Guide to Decision Making," to help IT organizations plan for that transformation.**

As recently as a few years ago, few IT organizations were concerned with application delivery. That has all changed. Application delivery is now a top-of-mind topic for virtually all IT organizations. As is described in this handbook, there are many factors that complicate the task of ensuring acceptable application performance. This includes the lack of visibility into application performance, the centralization of IT resources, the decentralization of employees and the complexity associated with the current generation of n-tier applications.

Some of the IT organizations that were interviewed for this handbook want to believe the challenges associated with application delivery are going away. They want to believe application developers will soon start to write more efficient applications and bandwidth costs will decrease to the point where they can afford to throw bandwidth at performance problems.

The complexity associated with application delivery will increase over the next few years.

That follows in part because, as explained in this handbook, the deployment of new application development paradigms such as SOA (Services Oriented Architecture), Rich Internet Applications and Web 2.0 will dramatically increase the difficulty of ensuring acceptable application performance. It also follows because of the increasing management complexity associated with the burgeoning deployment of the virtualization of IT resources (i.e., desktops, servers, storage, applications), the growing impact of wireless communications, the need to provide increasing levels of security, as well as emerging trends, such as storage optimization.

Instead of reaching a point where the challenges associated with application delivery are going away, we are just ending the first phase of a fundamental

transformation of the IT organization. At the beginning of this transformation, virtually all IT organizations were comprised of a myriad of stove-piped functions. By stove-piped what is meant is that these functions had few common goals, terminology, tools and processes. A major component of the transformation is that leading-edge IT organizations are now creating an environment that is characterized by this realization:

If you work in IT, you either develop applications, or you deliver applications.

Put another way, leading-edge companies are creating an IT organization that is comprised of two functions: application development and application delivery. Both of these functions must work holistically in order to ensure acceptable application performance.

This view of IT affects everything - including the organizational structure, management metrics, requisite processes, technologies and tools. While the transformation is indeed fundamental, it will not happen overnight. We have spent the last few years coming to understand the importance and difficulty associated with application delivery and to deploy a first generation of tools, typically in a stand-alone, tactical fashion. As we enter the next phase of application delivery, leading-edge IT organizations will develop plans for how they want to evolve from a

“This lack of emphasis on application performance in a WAN environment often results in the deployment of "chatty" applications.”

stove-piped IT infrastructure function to an integrated application delivery function.

Senior IT management needs to ensure that their organization evolves to where it looks at application delivery holistically and not just as an increasing number of stove-piped functions.

This transformation will not be easy in part because it crosses myriad organizational boundaries and involves rapidly changing technologies that have never before been developed by vendors, nor planned, designed, implemented or managed by IT organizations in a holistic fashion.

Successful application delivery requires the integration of tools and processes. One of the goals of this handbook is to help IT organizations plan for that transformation - hence the subtitle: A guide to decision making.

### The Application Development Process

In most situations, the focus of application development is on ensuring that

applications are developed on time, on budget, and with few security vulnerabilities. That narrow focus, combined with the fact that application development has historically been done over a high-speed, low-latency LAN, means that the impact of the WAN on the performance of the application is generally not known until after the application is fully developed and deployed.

In the majority of cases, there is at most a moderate emphasis during the design and development of an application on how well that application will run over a WAN.

This lack of emphasis on application performance in a WAN environment often results in the deployment of "chatty" applications, as shown in Figure 3.1.

A chatty application requires hundreds of application turns to complete a transaction. To understand the potential impact of a chatty protocol, assume a given transaction requires 200 application

turns. Further assume the latency on the LAN on which the application was developed was 1 millisecond, but that the round trip delay of the WAN where the application will be deployed is 100 milliseconds. For simplicity, the delay associated with the data transfer will be ignored and only the delay associated with the application turns will be calculated. In this case, the delay on the LAN is 200 milliseconds, which is not noticeable. However, the delay on the WAN is 20 seconds, which is very noticeable.

The preceding example demonstrates the need to be cognizant of the impact of the WAN on application performance during the application development lifecycle. In particular, it is important during application development to identify and eliminate any factor that could have a negative impact on application performance. This approach is far more effective than trying to implement a work-around after an application has been fully developed and deployed.

The preceding example also demonstrates the relationship between network delay and application delay.



FIGURE 3.1: Chatty Application

A relatively small increase in network delay can result in a very significant increase in application delay.

#### Web 2.0 Performance Issues

... [T]he movement to a Service-Oriented Architecture (SOA) based on the use of Web services-based applications is going to drastically complicate the task of ensuring acceptable application performance. The same is true for the movement to Web 2.0. In the case of Web 2.0, however, the problem is exacerbated because most IT organizations are not aware of the performance issues associated with Web 2.0.

As noted [previously], the existing network and application optimization solutions were designed to mitigate the performance impacts of large payloads and multiple application turns. Microprocessor vendors such as Intel and AMD continually deliver products that increase the computing power is available on the desktop. As a result, these products minimize the delays associated with client processing (Cc). This leaves just one element of the preceding model that has to be accounted for - server-side delay.

This is the critical performance bottleneck that has to be addressed in order for Web 2.0 applications to perform well.

The existing generation of network and application optimization solutions does not deal with a key requirement of Web 2.0 applications - the need to massively scale server performance.

The reason this is so critical is that unlike clients, servers suffer from scalability issues. In particular, servers have to support multiple users, and each concurrent user consumes some amount of server resources: CPU, memory, I/O. Chris Loosley highlighted the scalability issues associated with servers [in his white paper titled "Rich Internet Applications: Design, Measurement and Management Challenges" (2006)]. Loosley pointed out that activities such as catalog browsing are "relatively fast and efficient" activities that do not consume a lot of server resources. He contrasted that to an activity that required the server to update something, "such as clicking a button to add an item to a shopping cart." His paper points out that an activity like updating consumes significant server resources and that "the number of concurrent server transactions plays a critical role in determining server performance."

The CEO [of a Mobile Software enterprise who was interviewed for this book] addressed the issue of scalability when he stated that there is no better application framework than ASP.NET, but that ASP.NET

does make it very easy to develop applications that do not perform well. As [this CEO] sees it, IT organizations need to answer the question of "How will we scale Web 2.0 applications that have a rich amount of information from a dynamic database?" He said that a big part of the issue is because of the dynamic content associated with Web 2.0 applications, "caching is not caching - it is different for every single application that you work with". As a result, IT organizations need to answer questions such as: "When can I cache that data?" and "How do I keep that cache up to date?" He added that the best way to solve the Web 2.0 performance problems is to deploy intelligent tools.

The CTO [of a Business Intelligence enterprise who was also interviewed] pointed out that the most important server-side issue associated with traditional applications was providing page views, while with Web 2.0 applications, it is supporting API calls. He emphasized that "You cannot scale a Web site just by throwing servers at it. That buys you time, but it does not solve the problem." His recommendation was that IT organizations should make relatively modest investments in servers and make larger investments in tools to accelerate the performance of applications.

# VoIP Quality of Experience. It's Your Call.

## GAIN VISIBILITY INTO VOIP PERFORMANCE WITH THE NETQOS VOIP MONITOR

The NetQoS VoIP Monitor is a network-based call setup and call quality monitoring product that gives you the visibility to ensure a satisfactory Quality of Experience for VoIP users while ensuring network performance. It monitors the call quality, provides alerts on call performance problems, and isolates performance issues to speed troubleshooting and mean time to repair.

Integrated with the NetQoS Performance Center, it gives you the only complete solution for managing converged voice, video, and data applications on your enterprise network.



Make the connection between your users' Quality of Experience and your networks' Quality of Service with the **NetQoS VoIP Monitor**.


**SCHEDULE A DEMONSTRATION TODAY.** U.S. Toll-Free: (877) 835-9575 | [www.netqos.com/vm](http://www.netqos.com/vm)

United Kingdom: 44 (0) 118 929 8032

© 2008 NetQoS, Inc. All rights reserved.

NetQoS and the NetQoS logo, are registered trademarks of NetQoS, Inc. in the United States and in other countries.

**NetQoS**  
Performance First



# TECH BRIEFS:

Jim McQuaid and Brad Webster, NetQoS

UNDERSTANDING AND MONITORING  
ECHO CANCELLATION FOR OPTIMAL  
VOIP PERFORMANCE



Echo is a troubling problem for telecom professionals. Most of us have suffered through a telephone call where we had to try to talk with a lot of echo on the wire. It's very distracting.

Voice over IP does not create echo. However, due to the temporal aspect of echo, VoIP systems can and do increase the amount of echo heard during a telephone conversation. Echo cancellation is one of the more complex parameters that needs tuning to preserve VoIP call quality. IP networking specialists are increasingly finding that they must understand and monitor echo cancellation to manage their VoIP system.

Here, we present an overview of echo issues, provide a brief introduction to echo cancellation, and describe the echo metrics provided by the NetQoS® VoIP Monitor VoIP management solution.

### What Is Echo?

Echo is your voice coming back to you, as if you were repeating yourself. During a normal, two-person phone conversation, your voice is transmitted from your mouth to the ear of the person at the other end, and their voice is returned from their mouth to your ear. However, in any conversation, a certain amount of your own voice is also part of what you hear, whether you are talking face-to-face with someone who is sitting in your office, or

talking to someone on the phone. This experience of hearing your own voice is not echo. Commonly called "sidetone," it's a normal aspect of talking and listening. Your own voice becomes "echo" when it comes to your ear with a significant delay from the time you spoke. Sidetone is scarcely noticeable when the delay between your speaking and hearing is less than 25 milliseconds. Within that time window, the human brain does not perceive the sound as echo.

Echo, then, is very much a function of latency. Once you can hear your own voice more than 25 ms later, the possibility of perceiving it as echo arises. Twenty-five to 150 ms is a typical delay range for international telephone calls, which is why echo cancellation is necessary for such calls. Voice over IP calls don't actually create additional echo, but they also have a delay budget in the range of 150 ms to preserve audio quality, so VoIP systems commonly employ echo cancellation as well.

Among the various settings and parameters that need tuning to preserve optimal call quality in a VoIP system, echo cancellation is one of the more complex, least understood factors that IP networking specialists must address when

they encounter VoIP for the first time. To understand echo cancellation and the metrics associated with it, we need to look at some other aspects of echo that affect the implementation of echo cancellation on voice gateways.

### Echo Is Never Digital

Echo is always caused by the analog components in the telephony system. The digital stream of packets traveling in one direction of a VoIP call cannot "bleed into" the digital stream of packets in the other direction, nor are the packets played back at the receiving end of the call. The same is true for the digital parts of the Public Switched Telephone Network (PSTN): while the underlying electrical signals carrying the bits over the traditional switched telephone network are, indeed, analog, the corruption of those signals results in digital noise or other problems, but not in echo. Strictly speaking, echo is never caused by voice over IP. In fact, what happens is that the longer delays introduced by all voice over IP systems reveal echo that was imperceptible with the shorter delays of the PSTN. By delaying existing echo signals longer, the VoIP network causes them to fall outside that 25 ms window and become audible to us.

### What Causes Echo?

To reiterate what we've said so far, echo is the reflection or return of the speaker's voice to the speaker. It has an analog source, and it usually occurs at the far end of a conversation. Cisco Systems explains that the main two types of echo have different sources:

- Hybrid echo-Caused by an impedance mismatch in a hybrid circuit, such as a two-wire to four-wire interface, which allows the Tx signal to appear on the Rx path.
- Acoustic echo-Caused by poor insulation between the earpiece and the microphone in telephone handsets and hands-free devices.

At several places along a phone circuit, your voice can get into the return channel and come back to you. The first interface where echo may occur is at the transition between a 4-wire and a 2-wire interface. Analog telephone handsets are 2-wire devices. At some point in the path, perhaps in a local PBX, there is a hybrid interface that converts the network 4-wire interface to the 2-wire interface. Impedance mismatches here will reflect some of the energy back into the network, creating a potential source of hybrid echo. Another common source of echo is the basic hardware: the mouthpiece of the phone at the far end may be too close to the earpiece, or it may be poorly insulated, so that your voice is heard and forwarded on the same return channel as the one on which the person at the far end is speaking. Therefore, the analog phone itself is a possible source of acoustic echo. Even more suspect these days is the speaker phone function of the phone at the far end of the call. Speaker phones broadcast the voice and simultaneously listen to the voices of the speakers in the room. It is all too easy for speaker phones-

especially cheap ones-to send back some of the far end voice as part of what they are "hearing."

Delay is a necessary condition for echo, so it is rare for components that are close to the speaker-that is, on the speaker's side of the call-to cause echo. Even if part of the transmitted signal is being reflected back to the speaker by means of the return channel, the propagation delays are so brief that it will never be heard as echo. But several required components in every VoIP system exacerbate delay. The extra latency starts with the codec, which translates the analog signal into digital packets and places these packets on the wire. Latency is often increased by network components, such as routers, by geographical distance, and by jitter buffers in the IP phones. Any network congestion only makes it worse.

Because echo is rarely caused by a local component and has an analog source, the main suspects when an echo problem crops up are usually part of the "tail circuit" connecting the remote speaker to the PSTN. See Figure 1, below, for an illustration. The voice gateway device shown in the diagram allows analog phone calls from the PSTN to enter the IP network, and vice-versa.

### Echo Metrics and Cancellation

If echo usually occurs at the far end of a call, echo cancellation is ideally done at the far end of the call as well. However, it can be done at any analog boundary in the network, with varying degrees of success. In one sense, you cannot eliminate an echo originating at the far end of the circuit if you don't control the equipment at the far end. However, you can minimize it using echo cancellation (or ECAN) devices.

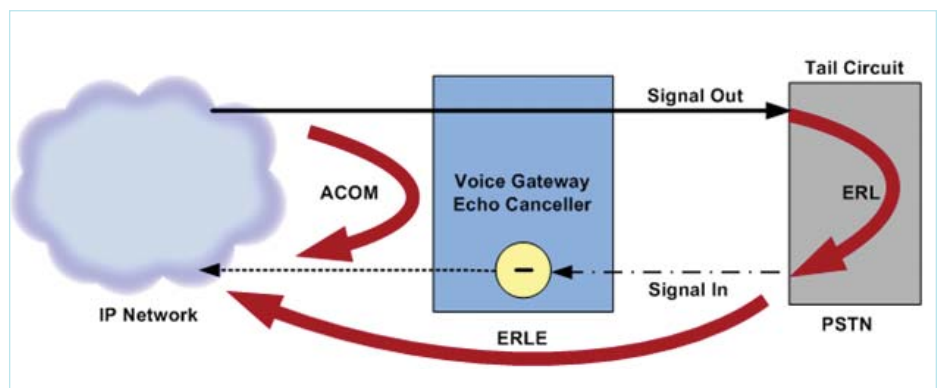


FIGURE 1. Echo Measurements & Echo Cancellation

// Not surprisingly, the louder the echo, the more distracting it is. Echo cancellation in effect consists of attenuating the amplitude of the echo part of the signal so that the echo is not easily heard. //

One place where telecom or network engineers try to minimize echo in a VoIP system is at the voice gateway connecting an IP network to the PSTN. Even though the echo is originating at the far end of the call, echo cancellation here can work, within limits.

An echo cancellation device operates by comparing the signal going into the tail circuit with the signal coming back out. Basically, the ECAN remembers the signal pattern of the signal entering the tail circuit and examines the data exiting the circuit to see if it contains this pattern. If I say, "Is that good?" and you reply, "Fine" the ECAN is remembering "Is that good?" and looking for it mixed in with "Fine." Two dimensions of the ECAN's work are important: echo strength, or volume, and echo delay.

### Echo Volume

Not surprisingly, the louder the echo, the more distracting it is. Echo cancellation in effect consists of attenuating the amplitude of the echo part of the signal so

the echo is not easily heard. As some point, the echo becomes so quiet that it disappears, as compared to the foreground volume of the call.

Volume is typically measured in decibels (dB). You've probably seen the charts that show a whisper is in the range of 20 dB while a jet engine or a rock band you don't like is around 120 dB.

Echo strength, which is equivalent to volume, is measured as echo return loss (ERL), which we discuss below. The ERL sets some boundaries on the ECAN's function because the echo must be weak enough to be distinguished from regular speech. In practice, this means that echo must be 6 dB quieter than the speech it appears alongside for the ECAN to be able to suppress it. If the echo is louder than this, it typically falls into the volume range of the actual replying speaker and cannot be safely removed without endangering the conversation contents. So, if an echo were actually present in a telephone conversation and so loud that the ECAN could not distinguish it from the

conversation, you would be more or less doomed to a very bad call with high levels of echo.

### Echo Delay Times

Delay is the other key dimension of echo cancellation. As the ECAN examines the signal returning from the tail circuit, looking for a pattern that matches the signal sent into the tail circuit, it applies a convergence time algorithm to find the echo portion. Basically, this means ECANs obey a time window that limits their operation. They look for the signal going into the tail circuit to be repeated in the signal coming back within a specific time, like 12 ms. (Typical values are from 8 to 64 ms.)

### ECAN Limitations

Because the ECAN is actually modeling the response of the tail circuit mathematically, it starts each conversation with no knowledge of which part of the signal is legitimate speech and which part is echo and has to build its model. This usually takes a few seconds. After that, the ECAN's ability to discern and remove echo reaches its operating state for that call, and most echoes fade.

If a signal included multiple reverberations at different delays, like ripples in a pond, the ECAN would see perhaps the first two, but miss the one that arrived after 12 ms had passed. Typically, the energy level of each successive reverberation is reduced, so the resulting echo would potentially be quite soft anyway.

One other condition that can exceed the ability of an ECAN is distortion. If the echo itself is so distorted that it no longer matches the pattern seen as the signal was sent into the tail circuit, the ECAN will not be able to recognize it as echo.

### ECAN Metrics

Taking into consideration the fact that echo cancellation is really all about echo suppression and not echo prevention, you've probably already deduced that echo is extremely common in any telephony environment. And as we mentioned above, it plagues VoIP systems due to their multiple sources of delay. It's therefore important to closely monitor the echo levels in a given VoIP system. Unless they're using a mobile phone and expecting slightly inferior service, users will complain stridently if echo becomes noticeable during their phone calls.

When monitoring echo levels in a phone system, or when troubleshooting a reported issue with echo, telecom professionals apply a well-known set of metrics that express the effectiveness of echo cancellation. These metrics also apply to VoIP, within limits.

### Echo Loss Metrics

Echo return loss, or ERL (see Figure 1), is a measurement applied to echo that measures the loss of volume between the original signal and the echo. In other words, an ERL of zero is the worst case; it means that the echo is fully as loud as the original signal. As the delay grows longer,

an ERL of up to 55 dB (and at least 6 dB) is necessary to soften the echo enough to avoid distraction. Applying this principle to VoIP, with long delays (up to 150 ms for acceptable call quality), the echo part of the signal needs to be at least 15 dB quieter than the original voice in order to avoid the perception of echo. So high ERL values are good. Note that ERL and related values mentioned in this document are all part of various ITU standards.

The amount of echo suppression is measured in the ERLE, or echo return loss enhancement, which expresses how much quieter the ECAN was able to make the echo, in dB. In other words, ERLE is a measure of what the actual echo canceller is able to accomplish. While ERL measures the "native" echo coming from the tail circuit or far end of the call, ERLE is the amount of additional echo attenuation the ECAN provides. Taken together, they exactly equal the ACOM value.

Another standard metric applied to echo cancellation, the ACOM value is the view of echo from the IP side of the echo canceller. As defined in ITU G.168, ACOM is the "combined" echo return loss through the system—the attenuation of echo from all possible means.

ACOM resembles ERL: it is a measure of the degree to which an echo signal has been attenuated. The difference is that ACOM is measured on "our side" of the ECAN device (as shown in Figure 1, above). Therefore, because it includes all sources of echo loss in each direction of the circuit, ACOM is the best gauge of echo strength. Like ERL, ACOM should be high.

### Signal In and Signal Out Metrics

The Signal In and Signal Out metrics are useful for testing a circuit to see whether it is introducing echo. Cisco Systems' echo testing procedures use these metrics to measure the effects of altering the signal strengths when tuning echo levels. Signal In is the audio signal traveling in the

**“Taking into consideration the fact that echo cancellation is really all about echo suppression and not echo prevention, you've probably already deduced that echo is extremely common in any telephony environment.”**

direction of the IP network, measured as it enters the ECAN from the tail circuit (shown as "Signal In" in Figure 1). It contains echo that needs to be canceled.

Signal Out is the audio signal coming out of the ECAN and going into the tail circuit (shown as "Signal Out" in Figure 1)-from the IP network to the PSTN. The Signal Out stream contains an estimation of the amount of echo in the audio stream. Both metrics are measured by the gateway on the PSTN side of the gateway's ECAN.

ERL is used along with Signal In and Signal Out to tune the echo canceller. Remember that ERL must be at least 6 dB to distinguish the echo portion of the signal from the voice itself. To enhance this difference and enable echo cancellation, the voice gateway ports perform "input gain" and "output attenuation" on the signals. Input gain is performed at "Signal In" in the diagram, before the gateway's ECAN sees the echo, and output attenuation is performed at Signal Out, after the gateway's ECAN has seen and cached the original signal. Thus, these metrics provide reference points for adjusting the overall strength of the signals to enable echo cancellation. They are used in echo troubleshooting to adjust the signal gain or attenuation performed by the gateway so that:

**Signal Out - Signal In > 6 dB.**

### Addressing Echo Problems in a VoIP System

The NetQoS VoIP Monitor product ensures the availability and performance of your voice over IP (VoIP) system by passively monitoring and reporting on VoIP call

setup and call quality metrics. NetQoS VoIP Monitor has several features that make it uniquely capable of measuring echo levels in voice over IP telephone calls, sending alerts when echo crosses a threshold, and helping track down the source of an echo problem.

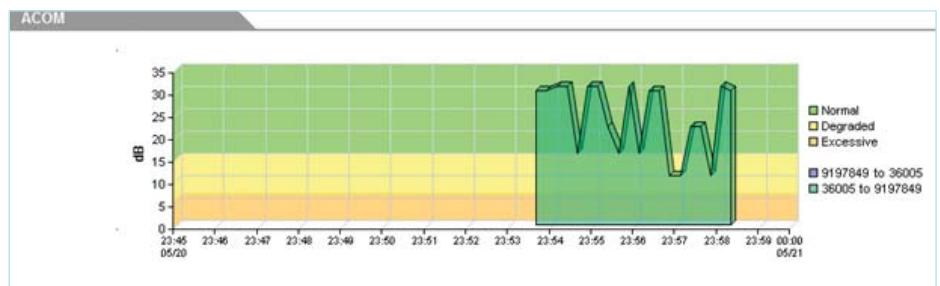
The VoIP Monitor system continually monitors both call setup performance and call quality. ACOM is reported for all call legs that include a voice gateway. One of the default call quality threshold settings instructs the system to raise alerts if ACOM values are too low. Like ERL, ACOM should be high; a VoIP Monitor alert is sent by default when ACOM drops to 15 dB. We stated earlier that ECANs use a limit of 6 dB to distinguish echo and avoid suppressing the actual contents of a conversation. If the ERL value is too low, the echo signal that returns to the gateway might be too loud, falling within 6 dB of the conversation signal. With the default threshold settings, the VoIP Monitor Management Console flags call quality as excessively bad when ACOM measurements, or the sum of ERL + ERLE, fall to 6 dB. ACOM measurements are available per voice gateway, so you can easily spot a gateway where excessive

echo is causing call quality to deteriorate.

In addition to threshold monitoring, NetQoS VoIP Monitor also reports ERL and ACOM values in real time for "watched" calls. The VoIP Monitor Call Watch feature collects additional diagnostic data from selected VoIP calls. During a Call Watch, VoIP Monitor actively gathers detailed quality metrics for all calls made to and from a selected IP phone by polling the phone and any associated gateway, if a call to the PSTN is watched. The collected data is presented in a series of charts, which are displayed and updated in real time, as the watched calls are in progress.

The Call Watch Report also includes Signal In and Signal Out metrics. In addition to contributing to the information-gathering phase of troubleshooting a call-quality problem, these values can help engineers tune their echo cancellers. In order to tune the ECAN, test signals are introduced into a voice gateway circuit from an IP phone. Gain and attenuation are applied through the command line at the incoming and outgoing gateway interfaces until the desired values of ERL and ERLE are obtained.

Cisco Systems' typical echo testing



**FIGURE 2.** Real-time ACOM measurements are compared to the Degraded and Excessive performance thresholds.

procedures expect you to gather these metrics by frequently running a command from the command-line interface during an active phone call. Now you can simply start up a VoIP Monitor Call Watch for a phone connected to the gateway under test, make a phone call to the PSTN to or from that phone, and take a look at the metrics as they are reported in real time. The signal levels are presented in a graph format that is continually updated during the tuning procedure.

## Summary

Echo is a difficult problem and not one that can be readily solved without access to the end-to-end circuit. Echo cancellation devices at the boundary of the VoIP system attempt to reduce echo that is of no concern in the PSTN so that it cannot be heard in the IP telephony environment. NetQoS VoIP Monitor provides a quick, overall rating of the success of this effort in the form of ACOM values for all gateway calls made on the monitored network, as well as some additional details, such as ERL, Signal In, and Signal Out, that are essential for troubleshooting echo problems.

## Helpful References

NetQoS VoIP Monitor Product Overview  
[http://www.netqos.com/solutions/voip\\_monitor/index.html](http://www.netqos.com/solutions/voip_monitor/index.html)

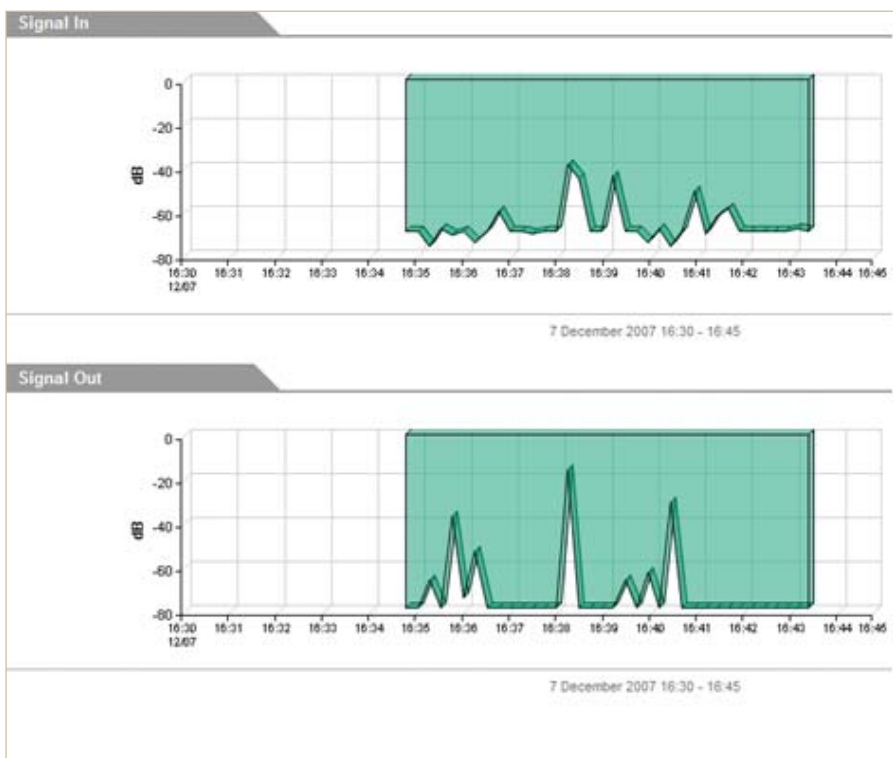
E-Book: Do You See What I'm Saying?: Managing VoIP Quality of Service on your Network  
<http://www.netqos.com/ebook/>

Cisco: Troubleshooting Echo Problems between IP Phones and IOS Gateways  
[http://www.cisco.com/en/US/tech/tk652/tk698/technologies\\_tech\\_note09186a0080149a1f.shtml](http://www.cisco.com/en/US/tech/tk652/tk698/technologies_tech_note09186a0080149a1f.shtml)

Cisco: Parameters important to troubleshoot echo  
[http://www.cisco.com/en/US/tech/tk652/tk698/technologies\\_tech\\_note09186a008019ab88.shtml#topic6](http://www.cisco.com/en/US/tech/tk652/tk698/technologies_tech_note09186a008019ab88.shtml#topic6)

Cisco: IP Telephony Advanced Topics and troubleshooting  
[http://www.cisco.com/web/CA/events/pdfs/1P\\_Telephony\\_Advanced\\_Topics\\_And\\_Troubleshooting.pdf](http://www.cisco.com/web/CA/events/pdfs/1P_Telephony_Advanced_Topics_And_Troubleshooting.pdf)

International Telecommunications Union: G.168: Digital network echo cancellers  
<http://www.itu.int/rec/T-REC-G.168-200408-S/en>



**FIGURE 3.** Signal In and Signal Out metrics are useful for tuning an echo canceller.

# WHITEPAPER

Ben Erwin, Technical Marketing Manager, NetQoS

## MANAGING THE PERFORMANCE OF FINANCIAL TRADING APPLICATIONS

With trillions of dollars traded annually on the NASDAQ alone, financial services companies are investing heavily in optimizing their electronic trading infrastructure and providing nearly instantaneous access to the latest market data.

The widespread adoption of electronic trading has spurred many global IT initiatives to reduce application latency to extremely low levels. Put simply, a one-millisecond advantage can be worth millions. Acquiring and sustaining this advantage has created a fiercely competitive landscape, where financial services companies must keep a watchful eye on their trading infrastructure.

Alongside competitive advantage, regulatory initiatives are also forcing trading services to provide market access in a timely manner. Now more than ever, financial services companies need the right solutions to manage the delivery of their trading applications.

The financial services industry has standardized on a communication protocol for financial trading applications called Financial Information eXchange (FIX).

For the trade desk, managing application delivery requires an understanding of the FIX protocol; however, many financial services companies are not equipped with

the right solutions to monitor and troubleshoot FIX applications. Solutions that provide visibility into network traffic, hop by hop latency, data center application performance, and retrospective analysis of FIX traffic can provide the needed insight. Without these solutions, financial services companies are left in the dark when they need to validate low-latency trade execution.

In this essay, we define the issues and outline the capabilities required to effectively manage the delivery of electronic trading applications and order management systems.

### Introduction

Trading activity between exchanges, brokerage firms, hedge funds, and other financial services companies is a core component of the world economy. With trillions of dollars traded annually on the NASDAQ alone, financial services companies are investing heavily in optimizing electronic trading applications and infrastructure, with the aim of providing nearly instantaneous access to the markets.

Put simply, when buying or selling electronically, a one-millisecond performance advantage can be worth millions. The need to acquire and sustain this advantage has created a fiercely competitive landscape, where financial services companies must frequently upgrade and monitor their order management system (OMS) and IT infrastructure. At the same time, regulatory initiatives like the Regulation National Market System (Reg NMS) in the US and the European Union's Markets in Financial Instruments Directive (MiFID) are also forcing trading services to handle vastly increased volumes of market data and still provide order execution to all clients in a timely manner.

Issues surrounding government compliance and the need to maintain a competitive advantage are driving more trading firms to favor automated (or "algorithmic") trading applications over phone calls and human decision-makers. This trend is evident on the New York Stock Exchange, where half the trades in 2007 were managed by software applications. The industry has even standardized on a communication protocol for financial trading applications called Financial Information eXchange (FIX).

The widespread adoption of electronic trading has spurred many IT initiatives to reduce application latency to extremely low levels. The London Stock Exchange

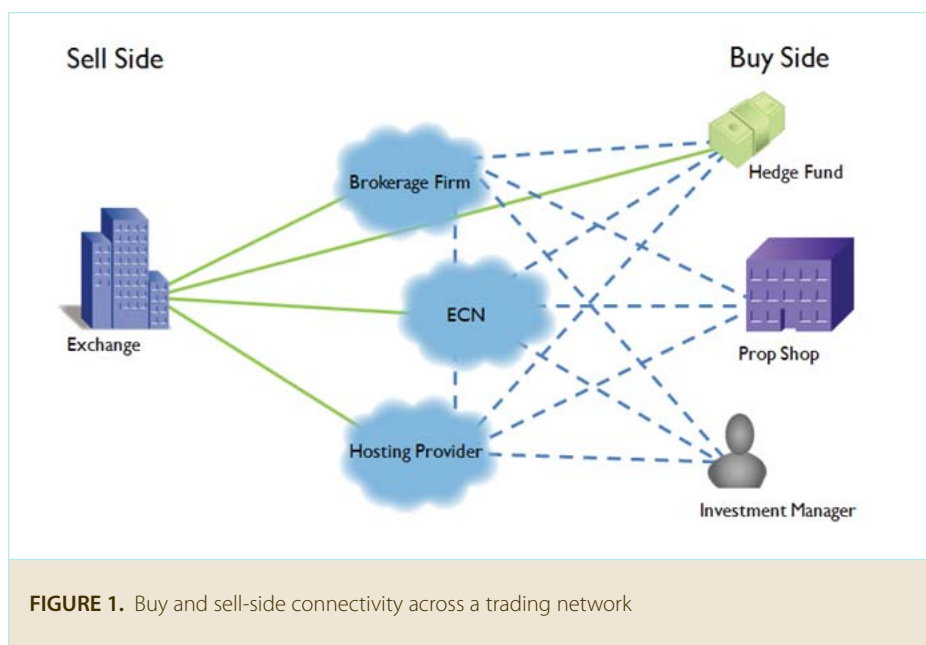
recently overhauled its trading platform to reduce latency from 110 milliseconds to 10 milliseconds. A US-based market exchange moved its data center across the country to shave latency in its order management system by 30 milliseconds. Hosting providers such as BT-Radianz and Savvis are getting into the race, providing buyers with direct market access (DMA) into data centers co-located with exchanges. DMA has grown in popularity over the last few years as more financial services companies show their willingness to invest heavily in low-latency architectures. Now more than ever, financial services companies need the right solutions to manage the delivery of their financial trading applications. For the trade desk, managing the delivery of applications requires an understanding of the FIX protocol; however, many financial services companies are not equipped with the right solutions to monitor and troubleshoot FIX applications. Solutions that provide visibility into network traffic, hop-by-hop latency, data center application performance, and retrospective analysis of FIX traffic can provide insight into application performance. Without these solutions, financial services companies are left in the dark when they urgently need to validate low-latency order execution. Latency is a critical metric, but a broader understanding is needed. With any networked application, a number of factors can affect performance, and in an order management system, poor performance can easily impact the bottom line. In this paper, we discuss managing

latency and performance in the fast-paced, constantly evolving sector of the financial services industry devoted to electronic trading. In this space, the ability to respond to the changes, keep ahead of competitors, and adhere to the rules is critical, and it requires the right tools. We'll begin by contextualizing our discussion, summarizing the key players in the industry and describing how they communicate.

### Buy Sides and Sell Sides

The financial services industry comprises two key groups: firms on the "buy side" and those on the "sell side." Buy sides are people or institutions that use market services; sell-side institutions provide market services. Buy sides typically connect into multiple sell sides, expecting lightning-fast trade execution and the

most up-to-date market data. Examples of buy-side firms include hedge funds, proprietary trading shops ("prop shops"), pension funds, and mutual funds. Examples of sell sides include brokerage firms, electronic communications networks (ECNs), and exchanges. You're probably familiar with buy-side institutions, but those on the sell side may be more obscure. An exchange (such as the NYSE or NASDAQ) is a marketplace for traders to buy and sell securities. Some buy sides will pay for direct market access, but, with the exception of large hedge funds, many buy side firms cannot afford it. The high cost of direct access provides opportunities for ECNs, brokerage firms, and hosting providers, who leverage economies of scale to provide affordable DMA into exchanges. An ECN is a trading network such as Bloomberg's TradeBook or Instinet





```
8=FIX.4.2|9=143|35=D|34=76|49=SGLM|56=SBI|
34=92|50=George|57=Joe|52=2000092604:39:59
|11=1|21=2|55=AMZN|54=1|38=5000|40=1|59=1|
10=044
```

FIGURE 2 - Sample FIX message

“The FIX protocol first surfaced in 1992 as an electronic communications method to be used between the financial services companies Fidelity Investments and Salomon Brothers.”

that matches buy and sell orders at specified prices. The example in Figure 1 shows communications between buy sides and sell sides.

Given the financial implications of timely trade execution, buy sides will select the sell-side providers with the fastest market access and/or guaranteed trade execution times. Therefore, sell sides (and buy sides paying for DMA) must keep a very close eye on the latency of their application services to ensure consistent service delivery and competitive advantage. The application services responsible for order execution leverage the FIX protocol; therefore, knowledge of FIX is important for understanding how quickly trades are being executed and where problems exist.

### The FIX Protocol

The FIX protocol first surfaced in 1992 as an electronic communications method to be used between the financial services companies Fidelity Investments and Salomon Brothers. FIX has undergone several revisions since its introduction, the latest being FIX v5.0, released at the end of 2006, and lately it has also seen explosive adoption rates along with the rise of algorithmic trading, investments in low-latency architectures, and the advent of regulatory compliance initiatives like RegNMS and MiFID. The FIX standard has several benefits, including real-time trade execution and vendor neutrality. FIX applications can also scale to high volumes of trade activity while achieving less than 10-millisecond delays. As a communication protocol, FIX is free, open, and widely used by software engineers who develop financial trading applications.

It can be thought of as a common electronic communication platform between financial services companies all over the world.

The FIX protocol standard is simple but verbose. Each FIX message contains information on the sending and receiving computer, ticker symbol, order type, number of shares, order time, and other relevant information to the trade. Each piece of information in a FIX message is represented as a series of tag-value pairs, as shown in Figure 2.

The challenges for financial institutions when managing the performance of FIX applications include minimizing server response time, responding to FIX protocol errors, and ensuring that trades are going through as requested. FIX applications must very rapidly interpret a high and variable volume of FIX messages, receive and process market data updates, and summarize thousands of these messages every minute, turning them into actionable information. IT organizations that manage trading networks and FIX applications need summarized performance reports, latency details, actionable alerts, and retrospective analysis capabilities to properly manage financial trading applications.

However, monitoring FIX applications alone is not enough to manage the delivery of trading services. These applications are primarily responsible for trade execution, but monitoring other

communications, like the critically important market data feeds that provide updated pricing and news that affects the markets, is also a necessity. In the next section of this paper, we describe all of the relevant IT components required for a complete view into trading application delivery.

### IT Technologies for Managing Trading Application Delivery

Managing the performance of financial trading applications requires visibility on multiple levels. Trading networks are a complex mesh of front- and back-office end points, and measuring latency between these end points can present a challenge. In addition, financial services companies may house thousands of servers in multiple data centers that are responsible for processing trades and sending instantaneous responses. They require highly scalable monitoring solutions to handle the trade traffic loads at these server clusters. Finally, with thousands of FIX messages traversing the network every minute, monitoring trade execution can be an overwhelming task.

Without deploying additional network infrastructure all over the globe to collect and analyze the data, how can financial services companies measure the delivery of their trading applications?

The answer is a multi-pronged strategy that includes source-to-destination monitoring, data center performance monitoring, and FIX forensic analysis.

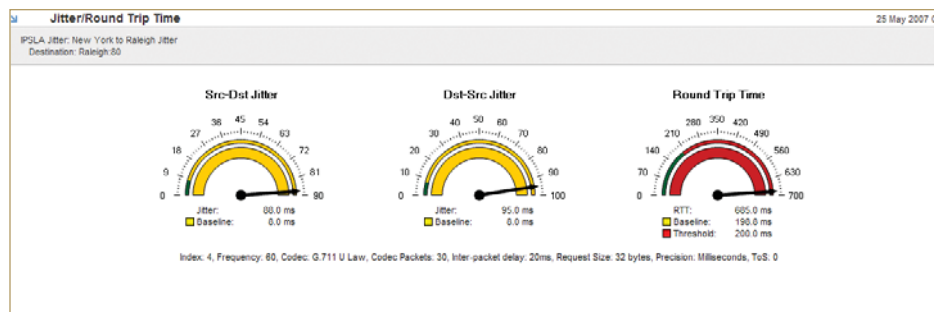


FIGURE 3 - Latency measurements from Cisco IP SLA

### Source-to-Destination Monitoring

Trading applications responsible for supplying the latest market updates communicate via one-way multicast blasts to multiple customers who are requesting the latest market prices. Determining hop-by-hop latency across a complex network mesh can be an expensive and time-consuming process without the appropriate monitoring capabilities. Cisco's IP SLA (Internet Protocol Service Level Agreement) technology embedded in Cisco routers is a quick method for determining hop-by-hop, one-way latency. IP SLA is an integrated component of Cisco IOS that can be leveraged in almost every Cisco routing environment to execute synthetic application transactions between source and destination end

points. Once a transaction has completed, IP SLA measures the latency of the transactions and makes the data available to third-party reporting solutions. An IP SLA-compatible reporting solution can extract these measurements from Cisco routers to provide actionable alerts and detailed analysis. See Figure 3 for an example of the types of metrics you can gather using IP SLA testing.

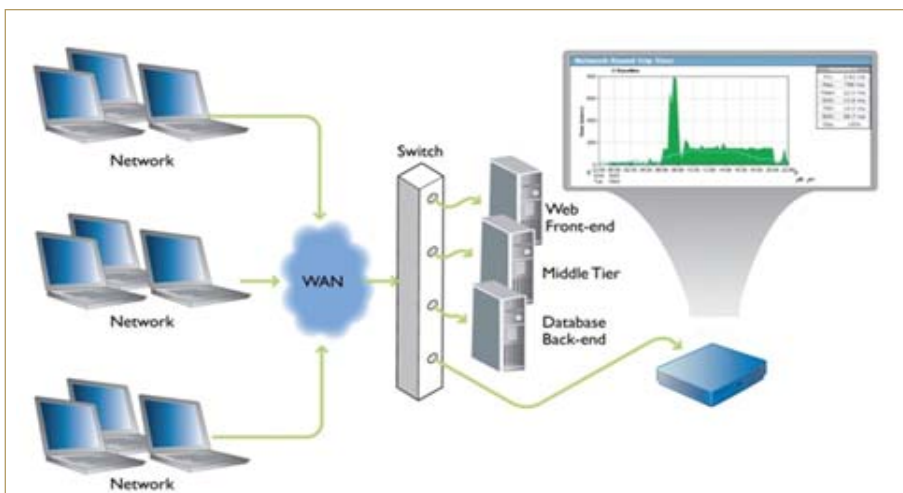
While latency is the most critical measurement for monitoring FIX application delivery, visibility into all of the traffic flows across a trading network is also necessary. Regulatory initiatives like RegNMS and MiFID have increased mandatory communications between financial services companies, resulting in

“ Without deploying additional network infrastructure all over the globe to collect and analyze the data, how can financial services companies measure the delivery of their trading applications? ”

“ Paired with a scalable collection infrastructure, Cisco NetFlow is an effective, straightforward way to achieve visibility into all of the applications that are sharing the trading network. ”

very high message volume across network links. And trading networks are not always dedicated to transporting market data; they sometimes serve as a path for other applications critical to a financial services enterprise. Traffic from business-critical applications, such as VoIP and email, may be found alongside trading applications on some network links and should be closely monitored as well. With a finite amount of network bandwidth, keeping an eye on how these applications consume network resources and comparing their consumption with that of trading applications is important for

maintaining the highest performance levels. Like Cisco IP SLA, Cisco NetFlow can be leveraged on a Cisco network infrastructure to provide details of traffic flows and traffic composition. NetFlow can export detailed records of every application transaction across the network to a third-party NetFlow collector. Paired with a scalable collection infrastructure, Cisco NetFlow is an effective, straightforward way to achieve visibility into all of the applications that are sharing the trading network.



## Data Center Performance Monitoring

A data center can create a latency bottleneck in a trading network. Financial services data centers are filled with complex, multi-tier server environments responsible for responding to order requests in sub-millisecond timeframes. Because these servers are critical to the delivery of trading services, installing third-party applications on them to monitor their performance is not advisable due to overhead, load, and security concerns. Rather, a less intrusive, passive approach to monitoring latency to and from servers via the switching infrastructure is preferred. Figure 4 provides a schematic architecture of such a nonintrusive approach, which monitors server traffic on a shared switch. In addition, to manage the overall delivery of trading applications proactively, performance monitoring solutions must be able to measure the latency of application flows and provide actionable alerts of performance degradations. These tools should also provide analyses that aid capacity planning and enable overall application performance to be assessed from the data center.

## FIX Forensics

As previously discussed, DMA connections are exploding in popularity, as buy sides continue to invest heavily in the lowest-latency connections into exchanges. Each of these DMA connections transports FIX application traffic between buy sides and sell sides to execute trades. Financial services companies need FIX-based monitoring solutions that can passively

FIGURE 4 - Passive monitoring of transactions between data center servers

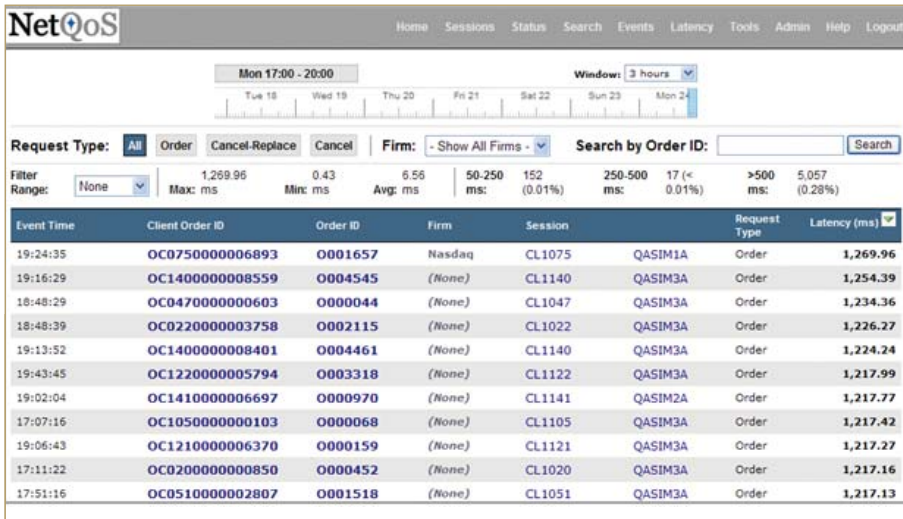


FIGURE 5 - FIX capture of performance outliers

monitor FIX transactions and record them for forensic analysis. A FIX monitoring solution should also be intelligent enough to match FIX requests and responses by presenting the data in a conversation format between clients and servers. This level of visibility allows both Trade Desk and IT personnel to immediately diagnose performance outliers and continue supporting millisecond trade execution times [Figure 5]. Actionable alerts and notifications are extremely valuable for the FIX trade desk to help minimize mean time to repair.

## Conclusion

Maintaining visibility and measuring latency throughout the trading network are must-haves for ensuring consistent service delivery for end users. Managing the delivery of trading applications can be a difficult process without a

comprehensive monitoring solution in place. Latency bottlenecks are extremely costly and can occur between any source and destination in the global client network, or within the data center infrastructure itself. The technologies currently available for gaining visibility into end-to-end performance and measuring FIX server latency include:

- Cisco IP SLA and NetFlow, for hop-by-hop source and destination monitoring across multicast networks. IP SLA provides latency metrics between end points, while NetFlow provides visibility into all application flows.
- Passive data center performance monitoring of the trading server infrastructure. Without the hassle and risks of

deploying software probes on servers, passive monitoring solutions can provide actionable information on application performance between servers in the data center and out to clients.

- FIX forensics monitoring of trading applications. This type of monitoring provides application-level, detailed information about every FIX session between buy sides and sell sides. A FIX monitoring solution should passively monitor trading applications for FIX server response times and record every transaction for any necessary trade verification or order troubleshooting.

Using the best-fit solution is a critical factor in maintaining a competitive edge in a cutthroat environment and retaining the trust of electronic trading clients. In addition to ensuring execution quality and minimizing latency, the right toolset will provide valuable assistance in infrastructure capacity planning and network management.

For more information on NetQoS Trade Monitor, visit [www.netqos.com/solutions/trade\\_monitor/](http://www.netqos.com/solutions/trade_monitor/).

# RESOURCES

## 2008-2009 CALENDAR OF RECREATIONAL NETWORK TRAFFIC MADNESS AND HOW TO USE NETFLOW TO ENSURE IT DOESN'T IMPACT NETWORK PERFORMANCE

Recreational traffic can have a major impact on the network. Numerous articles from a variety of sources examine the problems associated with the sporadic bursts of unauthorized network use from events such as March Madness (the annual National College Athletic Association men's basketball championship tournament, which takes place each March) and the NFL Super Bowl in January.

As the results of a recent NetQoS survey on recreational use of network resources show, the network performance problems associated with non-business usage of network resources is only getting worse, especially with the growing popularity of social media sites such as YouTube and MySpace. So, after years of helping customers prepare for periodic network traffic overload, NetQoS is now publishing a Calendar of Recreational Network Traffic Madness for the next six months.

This handy little calendar provides a month-by-month timeline of key events that can generate enough traffic to push many enterprise networks to the limits and adversely affect business-critical application performance.

### We also have a dynamic Google Calendar version of it up at

[www.networkperformancedaily.com/  
2008/01/network\\_recreational\\_traffic.html](http://www.networkperformancedaily.com/2008/01/network_recreational_traffic.html).

### December 2008

#### Monday, December 1

Cyber Monday  
Orange Bowl

#### Tuesday, December 2

Orange Bowl

#### Wednesday, December 3

Orange Bowl

#### Wednesday, December 31

Rose Bowl



## January

### Thursday, January 1

Rose Bowl  
The Queen's New Year's Honours List  
Announced

### Friday, January 2

Rose Bowl

### Sunday, January 4

Fiesta Bowl

### Monday, January 5

Fiesta Bowl

### Tuesday, January 6

Fiesta Bowl

### Wednesday, January 7

BCS Championship Game

### Thursday, January 8

BCS Championship Game

### Friday, January 9

BCS Championship Game

### Tuesday, January 20

Inauguration Day

### Friday, January 23

NHL All-Star Game

### Saturday, January 24

NHL All-Star Game

### Sunday, January 25

NHL All-Star Game

### Thursday, January 29

Super Bowl XLIII

### Friday, January 30

Super Bowl XLIII

### Saturday, January 31

Super Bowl XLIII

## February

### Sunday, February 1

Super Bowl XLIII

### Monday, February 2

Super Bowl XLIII

### Friday, February 6

NFL Pro Bowl

### Saturday, February 7

NFL Pro Bowl  
Rugby - Six Nations

### Sunday, February 8

NFL Pro Bowl  
Rugby - Six Nations

### Monday, February 9

NFL Pro Bowl

### Friday, February 13

Daytona 500

### Saturday, February 14

Daytona 500  
NBA All Star Game  
Rugby - Six Nations

### Sunday, February 15

Daytona 500  
NBA All Star Game  
Rugby - Six Nations

### Monday, February 16

Daytona 500  
NBA All Star Game

### Friday, February 27

Rugby - Six Nations

### Saturday, February 28

Rugby - Six Nations

## April

### Friday, April 3

NCAA Men's Final Four

### Saturday, April 4

NCAA Men's Final Four

### Sunday, April 5

NCAA Men's Final Four

### Monday, April 6

NCAA Men's Final Four  
Masters Tournament Augusta

### Tuesday, April 7

NCAA Men's Final Four  
Masters Tournament Augusta

### Wednesday, April 8

Masters Tournament Augusta

### Thursday, April 9

Masters Tournament Augusta

### Friday, April 10

Masters Tournament Augusta

### Saturday, April 11

Masters Tournament Augusta

### Sunday, April 12

Masters Tournament Augusta

## May

### Friday, May 22

Indy 500

### Saturday, May 23

Indy 500

### Sunday, May 24

Indy 500

### Monday, May 25

Indy 500

# Subscribe to the Performance Edge Journal.

**Free** subscription to qualified subscribers. Print issues are available in the US and Canada only. Online copies can be requested for all other locations at [www.performance-edge-journal.com](http://www.performance-edge-journal.com).

The Performance Edge Journal brings you in-depth analysis of network performance management issues. It provides technical analysis, tips, insight and reviews to thousands of Network and IT Managers to help them make informed decisions.

If you are responsible for **networking, voice, data, and video communications technologies** at enterprise, government or service provider organizations, this journal will bring you the content you need to do your job more effectively.

Subscribe today at [www.Performance-Edge-Journal.com](http://www.Performance-Edge-Journal.com).



- Technical Articles
- Industry Case Studies
- Analyst Commentary
- Best Practices
- Techniques
- Book Reviews
- And more.

[www.Performance-Edge-Journal.com](http://www.Performance-Edge-Journal.com)

EXCLUSIVE SUBSCRIPTION OFFER  
VISIT [WWW.PERFORMANCE-EDGE-JOURNAL.COM](http://WWW.PERFORMANCE-EDGE-JOURNAL.COM)

# PERFORMANCE EDGE JOURNAL

NETWORK PERFORMANCE  
MANAGEMENT SOLUTIONS  
FOR THE ENTERPRISE

**FREE** PUBLICATION FOR QUALIFIED  
NETWORKING, INFRASTRUCTURE  
AND IT PROFESSIONALS.

**To subscribe: Visit [www.Performance-Edge-Journal.com](http://www.Performance-Edge-Journal.com)**

PERFORMANCE EDGE  
JOURNAL

Attn: Marketing  
5001 Plaza On The Lake  
Suite 200  
Austin, TX 78746