

Managing Application Performance by Understanding Applications

By Jim Metzler

SPONSORED BY



Managing Application Performance by Understanding Applications

Introduction

One of the key characteristics of application development is that it is typically focused on ensuring that applications are developed on time, on budget, and with few security vulnerabilities. Another key characteristic of application development is that it is virtually always done over a high-speed, low-latency LAN. The combination of these two characteristics means that the impact of the WAN on the performance of the application is generally neither planned for nor known until after the application is fully developed and deployed.

In most organizations, if an application does not perform well it is assumed that the network is at fault. As a result of this assumption, typically the first step in identifying the cause of poor application performance is for the network group to prove that it is not the network. This defensive approach to problem identification often creates a strained relationship between the application development groups and the network organization.

A combination of factors is causing IT organizations to take application delivery more seriously. One of these factors is that with an ever increasing frequency organizations are automating their key business processes. As a result of this automation, if the applications that support these processes are not running well, the business processes themselves are not performing well. Another factor that is causing IT organizations to take application delivery more seriously is that ensuring acceptable application performance is continually becoming more difficult.

The goals of this white paper are to:

- Identify some of the key issues that make ensuring acceptable application performance so difficult.
- Describe how IT organizations can use WAN emulation tools to improve the performance of applications, plan for change and improve the relationship between the application groups and the network group.

To achieve these goals, three IT professionals were interviewed. One of the interviewees is the manager of technology, architecture and engineering for a large, international law firm. Another interviewee is a consultant who is responsible for application testing at a Fortune 500 pharmaceutical company, and the third is the lead performance engineer for a Fortune 500 financial institution. Throughout this white paper, the interviewees will be referred to as The Technology Manager, The Consultant, and The Performance Engineer.

IT INNOVATION REPORT

Published By

Kubernan
www.Kubernan.com

Cofounders

Jim Metzler
jim@ashtonmetzler.com

Steven Taylor
taylor@webtutorials.com

Design/Layout Artist

Debi Vozikis

Copyright © 2008

Kubernan

For Editorial and Sponsorship Information

Contact Jim Metzler
or Steven Taylor

Kubernan is an analyst
and consulting joint
venture of Steven Taylor
and Jim Metzler.

Professional Opinions Disclaimer

All information presented and opinions expressed in this IT Innovation Report represent the current opinions of the author(s) based on professional judgment and best available information at the time of the presentation. Consequently, the information is subject to change, and no liability for advice presented is assumed. Ultimate responsibility for choice of appropriate solutions remains with the reader.

The Requirement to Develop Efficient Applications

As mentioned in the introduction, applications have historically been developed over a LAN and little if any attention has been paid to how well the application would run over the WAN. One reason that an application performs well over a LAN and does not perform well over a WAN is the use of a chatty protocol as depicted in Figure 1.

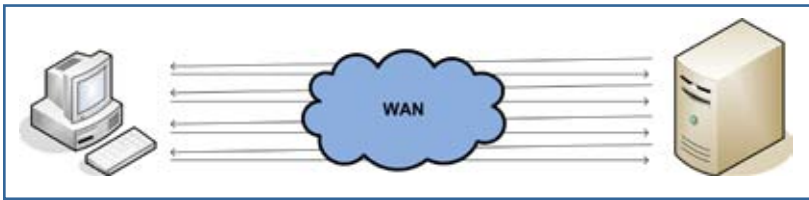


Figure 1: Chatty Protocol

A chatty protocol requires hundreds of application turns to complete a transaction. To exemplify the impact of a chatty protocol assume that a given transaction requires 200 application turns. Further assume that the latency on the LAN on which the application was developed was 1 millisecond, but that the round trip delay of the WAN on which 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 Technology Manager stated that in his organization chatty applications are pretty common and that the chattiness of these applications typically leads to long response times. He added that in some cases the response time is so long that the application becomes unusable and the company is forced to host a copy of the application in each of their twenty offices around the world. According to The Technology Manager, "Distributing the application often resolves the performance problems. However, it

significantly increases both the cost and the complexity of supporting the application."

The Consultant agreed with The Technology Manager relative to the prevalence of chatty applications and stated that in some instances the problem is so bad that "the application just can not be deployed." He added that in many cases there is not an easy way to improve application performance and that "there is just so much that you can do with caching ¹."

The discussion of chatty protocols demonstrates the need to be aware 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. The Technology Manager highlighted the importance of this approach when he stated, "It is much better to plan for the performance of an application during development than to scramble around after we develop it."

***"Half of the applications that we either develop or acquire are WAN vicious."
-- The Technology Manager***

The discussion of chatty protocols also demonstrates the relationship between network delay and application delay. As pointed out by the interviewees, a relatively small increase in network delay can

result a very significant increase in application delay. For example, The Technology Manager stated that application performance does not increase linearly as network delay increases. He added that in many cases network delay of 20 ms or less has minimal impact on application performance, but as network delay increases past 20 ms that the application delay "shoots up". The Performance Engineer summarized the experience of all of the interviewees when he stated, "An application that will work fine with 1 ms. of latency will sometimes run like a dog when there is 80 ms. of latency."

¹ Caching refers to storing a copy of the data close to the user to reduce the time it takes to access the data.

Other Complicating Factors

The development of applications that are not designed to run efficiently over the WAN is a major cause of poor application performance. However, as will be described in this section, there are a number of other factors that also cause poor application performance.

Server Consolidation

Many companies either already have, or are in the process of consolidating servers out of branch offices and into centralized data centers. This consolidation typically reduces cost and enables IT organizations to have better control over the company's data and hence allows the company to comply with governmental regulations such as the Sarbanes-Oxley Act (SOX).

Server consolidation typically results in protocols such as CIFS (Common Internet File System) running over the WAN. CIFS, which was designed to run over a LAN, is a chatty protocol. In particular, the way that CIFS works is that it decomposes all files into smaller blocks prior to transmitting them. The server sends each of these data blocks to the client where it is verified and an acknowledgement is sent back to the server. The server must wait for an acknowledgement prior to sending the next data block. As a result, similar to the example in the preceding section, opening a file that would take a fraction of a second before consolidating servers would take tens of seconds after the servers have been consolidated.

Data Center Consolidation

In addition to consolidating servers out of branch offices and into centralized data centers, many companies are also reducing the number of data centers that they support worldwide. HP, for example, recently announced that it was reducing the number of data centers it supports from 85 down to six². Many companies are also adopting a *single-hosting* model whereby users from all over the

globe transit the WAN to access an application that the company hosts in just one of its data centers.

"We are in the process of a large data center consolidation project. We are using WAN emulation to determine the location of the new data centers and to scale the infrastructure to accommodate the change."

-- The Technology Manager

Decentralized Work Force

At the same time that companies are centralizing their IT resources, they are typically decentralizing their work force. For example, as recently as a few years ago the vast majority of employees worked in a headquarters facility and accessed applications over a high-speed, low-latency LAN. The majority of companies, however, have moved away from this business model. Today there are over 6 million branch offices and 30 million branch office workers in the U.S.³ On a worldwide basis, there are over 23 million branch offices and 50 million branch office workers.

Branch office workers need access to the same applications as do workers in a headquarters facility. However, the combination of consolidating servers into centralized data centers while simultaneously decentralizing the work force means that the vast majority of workers now access applications over a WAN instead of a LAN. The fact that there is a movement both to consolidate data centers and to move to a single-hosting model for applications has the affect of increasing the distance between remote users and the applications they need to access. This increased distance translates into additional WAN latency, jitter and packet loss. The impact of increasing the distance between the user and the application is often not well understood. For example, The Performance Engineer pointed out that, "In most situations the application developers do not

² Hewlett-Packard picks Austin for two data centers <http://www.statesman.com/business/content/business/stories/other/05/18hp.html>

³ Addressing Operational Efficiencies in Branch Offices, IDC, May 2006

understand the impact on application performance of having the users be remote from the servers.”

Globalization

Over the last few years there has been an accelerating trend to distribute key business functions (e.g., R&D, marketing, manufacturing and customer service) around the country, and in some cases around the world. There are many factors driving this trend. One of these factors is the ongoing merger and acquisition activity. Another factor is that driven by a combination of logistical and economic necessity as well as legal requirements, many enterprises seek to have a presence close to their customers, suppliers and distributors.

Combining globalization with the trends discussed above results in an even longer WAN link, and hence more WAN latency, between the remote users and the applications they need to access.

Voice over IP (VoIP)

A number of studies ⁴ have shown that Voice over IP (VoIP) has begun to experience wide spread deployment. While VoIP provides a variety of strategic and tactical benefits, it does pose some unique challenges. For example, based on their experience over the last century with circuit-switched public telephony, users have come to expect 100% voice availability, fast call set-up and excellent quality. However, VoIP is very sensitive to network parameters such as delay, jitter and packet loss. As a result, when run over a packet network, voice does not always perform as well as it does when run on a circuit-switched network.

WAN Optimization

As previously noted, it is far more efficient to develop an application than performs well over a WAN than it is to implement a workaround after an application has been fully

developed and deployed. That being said, the last few years has seen the development of a class of device that is intended to compensate for the impact that the WAN has on application performance.

These devices are often referred to as WAN Optimization Controllers (WOCs) and they provide a wide variety of WAN optimization functionality including data compression, caching, TCP flow control, read-ahead and spoofing. One of the challenges facing IT organizations that are evaluating these devices is understanding the reasons why an application is performing poorly and then determining what WAN optimization functionality, if any, would improve application performance. For example, the performance of a chatty application would be improved by a WOC that did a good job of read-ahead and spoofing, but would not be improved by compression. Analogously, the performance of a bulk file applications such as data replication would be improved by a WOC that did a good job of compression, but would not be improved by read-ahead and spoofing.

SOA

The movement to a Service-Oriented Architecture (SOA) based on the use of Web services-based applications represents the next step in the development of distributed computing and will significantly increase the difficulty of ensuring acceptable application performance. To put that in context, in the typical n-tier application that is commonly deployed today the WAN impacts performance in just one place. That place is the link between the user's browser and the Web or application server. In a Web services-based application, the Web services that comprise the application typically run on servers that are housed within multiple data centers. As a result of housing the Web services in multiple data centers, the WAN impacts multiple traffic flows and hence has a greater overall impact on the performance of a Web services-based application than it does on the performance of traditional n-tier applications.

⁴ 2005/2006 VoIP State of the Market Report, Steven Taylor, www.webtorials.com

The Need for Effective Tools

IT organizations will not be regarded as successful if they do not develop an effective response to the issues raised in the preceding section. In particular, IT organizations must implement techniques during application development to ensure that when the application is deployed over a WAN that it will perform well. IT organizations must also develop the ability to understand in advance the impact of making a change such as deploying an application like VoIP, implementing new functionality such as WAN optimization, or making a major infrastructure modification such as consolidating data centers.

Relative to the topic of ensuring acceptable application performance, The Technology Manager stated that in the past 18 months his organization “found religion”. As previously noted, in that timeframe his organization has adopted the approach that it is better to plan for acceptable application performance during the application development phase than it is to “scramble around” to try to fix the performance problem once a poorly designed application has been deployed.

Over 200 IT professionals were recently asked, “Which of the following describes your company’s interest in a tool that can be used to test application performance throughout the application lifecycle – from application design through ongoing management?”⁵ The survey respondents were allowed to indicate multiple answers. Their responses are depicted in Table 1.

The fact that almost three quarters of the survey respondents indicated that such a tool would make a significant improvement to their ability to manage application performance highlights the widespread need for such a tool.

One class of tool that can be used to test application performance throughout the application lifecycle is a WAN emulation tool. These tools are typically used during application development and quality assurance (QA) and serve to mimic the performance characteristics of the WAN; e.g., delay, jitter, packet loss. One of the primary benefits of these tools is that application developers and QA engineers can use them to quantify the impact of the WAN on the performance of the application under development, while there is still time to modify the application. One of the secondary benefits of using WAN emulation tools is that over time the application development groups come to understand how to write applications that perform well over the WAN.

The Technology Manager said that as part of their recent focus on ensuring that applications will run well over the WAN, they now use a WAN emulation tool during application development. He added that they also used this tool to plan for the consolidation of servers out of their branch offices and into a set of centralized data centers. In particular, they used the WAN emulation tool to choose locations for their data centers that would result in acceptable application performance. In a similar fashion, The Consultant stated that they used a WAN emulation tool as

Response	Percentage of Respondents
If the tool worked well it would make a significant improvement to our ability to manage application performance	71%
Our operations groups lack the application specific skills to use a tool like this	17%
Tools like this tend to be too difficult to use, particularly during application development	13%
Our applications developers would be resistant to using such a tool	11%
The output of tools like this is generally not that helpful	9%

Table 1: Interest in an Application Lifecycle Management Tool

⁵ *The 2008 Application Delivery Handbook*, <http://www.webtorials.com/abstracts/Kubernan2008handbook.htm>

part of the process that they went through to consolidate data centers. He said that the quantitative data produced by the tool was very helpful because the overall process of consolidating data centers was “just a tad political” and the tool helped to quantify how the performance of the company’s key applications would change based on where the data centers were located.

The Consultant stated that “One of the real values of using a WAN emulation tool is that it helps to set user expectations in advance of deploying the application.” For example, his organization often uses WAN emulation during application development to predict application performance. They also get the end user organizations involved during the application development phase so that these organizations can look at the results produced by the WAN emulation tool and sign off (or not) on the acceptability of the performance. He added that the application development teams do not have to use WAN emulation during application development. However, he pointed out that his company recently deployed a high visibility application without using WAN emulation and that the application performance was miserable. That failure has made the application development teams notably more receptive to using a WAN emulation tool.

Whenever an IT organization is considering implementing a tool of this type it is important to choose a tool that is not overly complex. That is an issue with any type of management tool but it is a particular issue with a tool that is based on modeling or simulation. For example, when performing modeling or simulation, 80% of the insight can typically be provided while only incurring 20% of the complexity. However, in order to add additional insight requires the tool to become very complex and usually that requires a level of granular input that either does not exist or is incredibly time consuming to create.

“The application groups that use WAN emulation during development always come back and use it again.”

The data in Table 1 indicates that IT professionals are well aware of the fact that many of these tools are unacceptably complex. In particular, while the survey respondents indicated a very strong interest in these tools, thirty percent of the survey respondents indicated that tools like this tend to be difficult to use during application development or that their operations group would not have the skills necessary to use a tool like this. As a general rule, a tool that is unduly complex is of no use to an IT organization.

The Performance Engineer highlighted the importance of simplicity when he stated that, “One of the most important characteristics of a WAN emulation tool is simplicity. I want it to be simple enough that support people can use it without my help.” He added that, “It is important that the WAN emulation tool does not require me to have to install and configure extra routers just to use the tool.” The Consultant agreed with The Performance Engineer and said that, “Ease of use is a major consideration when choosing a WAN emulation tool. I want a variety of people to be able to be able to use the tool.”

The Technology Manager agreed that simplicity is important and stated that “When choosing a WAN emulation tool it is also important to choose a tool that has the ability to automatically analyze the existing network.” He also emphasized the importance of choosing a tool that supports effective reporting.

Bridging the Organizational Divide

As noted in the introduction, somewhat of a strained relationship often exists between the application development groups and the network organization. To gain insight into that relationship, two hundred IT professionals were asked, “How would you characterize the current relationship between your company’s application development organization and the network organization?”⁶ Their responses are depicted in Table 2.

⁶ *The 2008 Application Delivery Handbook*, <http://www.webtorials.com/abstracts/Kubernan2008handbook.htm>

Response	Percentage of Respondents
Highly Adversarial	0%
Moderately Adversarial	8%
Slightly Adversarial	17%
Neutral	33%
Slightly Cooperative	13%
Moderately Cooperative	25%
Highly Cooperative	4%

Table 2: Relationship Between Organizations

The data in Table 2 reflects a classic good news/bad news situation. The good news is that none of the survey respondents indicated that there was a highly adversarial relationship between their company's application development organization and the company's network organization. Part of the bad news is that only a very small percentage of the survey respondents indicated that there was a highly cooperative relationship between these two organizations. Another part of the bad news is that one quarter of the survey respondents indicated that there was some form of adversarial relationship between their company's application development organization and the company's network organization.

In a different survey, the survey respondents were asked to choose the two most significant impediments to effective application delivery ⁷. Table 7.3 shows the answers that received the highest percentage of responses.

Answer	Percentage of Companies
Our processes are inadequate	39.6%
The difficulty in explaining the causes of application degradation and getting any real buy-in	33.6%
Our tools are inadequate	31.5%
The application development group and the rest of IT have adversarial relations.	24.2%

Table 3: Impediments to Effective Application Delivery

The data in table 3 supports the preceding statement about the extent of the adversarial relationship between the typical application development group and the rest of IT. This data also highlights the fact that the adversarial nature of this relationship has the affect of making application delivery more difficult.

One of the reasons for the strained relationship between the application groups and the network group is that most application development groups have little knowledge of networks and most network organizations have little knowledge of applications. The use of a WAN emulation tool can help to bridge the knowledge gap between these two organizations. For example, as pointed out by The Performance Engineer, "The use of WAN emulations gives the applications developers a head-up that delay is a fundamental characteristic of the network and that we need to work with it and not around it." He went on to say that, "Application developers tend to have a 'show me' attitude. If you do not show them tangible evidence of where the problem is, they do not believe you."

The Consultant pointed out that sometimes the use of WAN emulation is not an option when he said that, "In many cases, there is an FDA requirements that mandates that we must test an application before it goes live." He agreed with The Performance Engineer that the use of a WAN emulation tool can help to bridge the knowledge gap between the application groups and the network group and added that, "The earlier in the lifecycle that you use a WAN emulation tool the better. They can help to identify a down and dirty architecture problem."

Summary and Call to Action

During the dot com era IT organizations were viewed as valuable due to the naïve notion that IT would enable companies to quickly and easily transform their business models. Shortly after the dot com implosion, IT was regarded merely as a utility, similar in importance to power and water.

⁷ The 2008 Application Delivery Handbook, <http://www.webtutorials.com/abstracts/Kubernan2008handbook.htm>

The good news is that IT is important again. The even better news is that IT is important again because of the support it provides for a wide variety of business initiatives, many of which were discussed in this white paper. Those initiatives include supporting data center consolidation, enabling a decentralized work force, as well as deploying new applications (VoIP) and new application architectures such as SOA. As a result of the breadth of business initiatives that IT is supporting, IT's importance seems to be more sustainable than it was during the dot com era.

In order to be successful in this new era of IT importance, IT organizations must avoid acquiring or developing applications that perform poorly when run over a WAN. In particular, IT organizations must adopt the approach that was described by The Technology Manager whose organization has decided that it is notably better to use a WAN emulation tool in order to plan for acceptable application performance during the application development phase than it is to scramble around to try to fix the performance problem once a poorly designed application has been deployed. Success, however, also requires that IT organizations do a better job of planning for major changes, such as consolidating servers into centralized data centers or consolidating data centers.

As demonstrated in this white paper, WAN emulation tools can help IT organizations achieve these goals. In order to start the process of deploying WAN emulation, IT organizations need to identify the characteristics that they are looking for in a WAN emulation tool (i.e., ease of use) and use those characteristics to choose the appropriate tool to meet their needs. The next step is to both create a compelling business case that quantifies the benefits of WAN emulation and to use this business case to get management buy-in. The final steps are to implement the tool and then market its use widely inside the organization in order to ensure that the organization gets the maximum benefit out of the tool.

About Shunra

Shunra Software is the market leader in WAN emulation solutions for application performance testing throughout the entire application development lifecycle. Shunra's Virtual Enterprise WAN Emulation Solution is an award-winning suite of solutions that enable development, quality assurance, pre-deployment and operations teams to create an exact replica of their production environment or design what-if network scenarios; allowing them to predict exactly how applications or infrastructure changes will perform in any networked environment—before rollout. Shunra's solutions and services are typically deployed in projects including data center consolidation; application performance readiness testing; VoIP testing; website testing and WAN optimization technology testing.

Shunra's solutions and services are used by thousands of customers worldwide. Clients include 3M, Boeing, Cisco, Dow Chemical, EMC, FedEx, General Motors, JPMorgan Chase, Kelly Services, Merrill Lynch, Motorola, Nestle, Pitney Bowes, and Vodafone. Shunra has offices in New York City, Philadelphia, Israel, and the UK. Shunra is also supported through a global network of channel partners. For more information on Shunra, please visit www.shunra.com.