A LAN Perspective on Wireless WAN Requirements

Dr. Jim Metzler

Ashton, Metzler & Associates

Leverage Technology & Talent for Success

1.0 Executive Summary

With the advent of the mobile workforce, one of the challenges facing IT organizations is to be able to take key business applications that were developed in a LAN environment and have them perform at an acceptable level in a WAN environment in general, and in a wireless WAN environment in particular. This challenge is made more difficult due to the fact that today none of the mobile operators offers a wireless WAN service that has performance characteristics (i.e., latency and throughput) that are anywhere close to the performance characteristics of a LAN.

With that challenge in mind, the goal of this paper is to provide a LAN perspective on wireless WAN requirements. In order to achieve that goal, this white paper will describe the complexity associated with enterprise applications and will present rules of thumb relative to the requirements of a number of representative applications. The paper concludes with the observation that **in order to support common enterprise applications, wireless WAN services need to have average sustainable latency of 150 ms or less, and an average sustainable send/receive throughput in excess of 300 Kbps.**

2.0 Introduction

The typical IT function has historically been comprised of numerous disparate organizations. However, within many companies the IT organizational structure is changing, albeit slowly. Inside of a number of IT functions there is the formation of two distinct organizations: the applications and the infrastructure organizations.

The applications organization focuses on acquiring, developing, and maintaining the business related applications that are used by the company's employees, partners, and customers. An Enterprise Resource Planning (ERP) application is an example of the type of application that is controlled by the applications organization. Note that **enterprise applications such as ERP are typically developed in a LAN environment.**

The infrastructure organization is responsible for planning, designing, implementing, and managing the company's Local Area Networks (LANs), Wide Area Networks (WANs), Storage Area Networks (SANs), Data Centers, and security. The infrastructure deployed by this organization needs to support all of the company's applications.

One of the challenges facing the infrastructure organization is being able to take all of the key business applications that were developed in a LAN environment, and have them perform at an acceptable level in a WAN environment in general, and in a wireless WAN environment in particular. With that challenge in mind, the goal of this paper is to provide a LAN perspective on wireless WAN requirements. In order to achieve this goal, this white paper will describe the complexity associated with enterprise applications and will present rules of thumb relative to the requirements of a number of representative applications.

The reason that it is important to understand the complexity that is associated with enterprise applications is because the complexity of a system has a significant impact on how the system needs to be designed. In particular, a complex system has many interconnected components that can impact the performance of the overall system. As will be shown in Section 6 of this white paper, **if any component of a complex application is under-provisioned (i.e., the WAN), it can have a significant negative impact on the performance of the overall system.**

3.0 Applications Management

In September 2003, Ashton, Metzler & Associates conducted research to quantify the role of applications management. One of the key results of that research is that applications management has become one of the most important components of network and systems management. Another major result of this market research is that **applications management is the most difficult component of network and systems management**.

In August 2003, Network World and Packeteer published a study entitled "Applications Performance Market Study". The study represents a survey given to 190 IT professionals who work in a wide range of companies, and who are closely involved with the management and operations of their company's networks and information systems.

Some of the key results of the study are the:

Extent of Application Performance Degradation

The majority of companies in general, and 85% of large companies (those with annual revenues in excess of one billion dollars) in particular, have experienced incidents of significant application performance degradation. In addition, the majority of large companies believes that the number of application slowdowns has increased in the last year. The two primary causes of application degradation are the architecture of the applications and the performance of the company's WAN.

Impact of Application Performance Degradation

The area of business in which applications performance degradation has the highest impact is lost employee productivity. However, there is also a significant negative impact on both customer service and revenues. Application degradation also leads to some employees refusing to use the application.

In addition, the majority of companies have either delayed launching new applications, or extended the rollout of new applications due to concerns about network application performance. The application that is delayed the most frequently is Enterprise Resource Planning (ERP).

Response to Application Performance Degradation

As mentioned, the two primary causes of application degradation are the architecture of the applications and the performance of the company's WAN. However, few companies intend to restrict applications as a means of ensuring application performance. Instead, the most common technique that companies plan to use in order to ensure application performance is to add bandwidth.

As mentioned in the introduction to this white paper, it is critical to properly provision each component of a complex system. Hence, in many if not most cases, **part of the proper response to application performance degradation is the addition of bandwidth with appropriate throughput and latency characteristics, as well as QoS functionality**.

4.0 The Wireless WAN Environment

A recent study by the Yankee Group[1] described the dynamics of the wireless WAN environment. Throughout this white paper, that study will be referred to as The Yankee Study.

The Yankee Study highlights the issues that limit the ability of enterprise IT organization to make additions of wireless WAN bandwidth with the appropriate characteristics and functionality. According to The Yankee Study, there are a number of applications that enterprises want to run over wireless WANs. Those applications include:

- Remote access to specific corporate data or applications
- Wireless email and calendar access
- Work force automation
- Browsing and Web services
- Customer facing applications

However, The Yankee Study also points out that the current commercially available wireless WAN data networks provide bandwidth of 100 Kbps or less and latency or 200 ms or more. As will be shown in the subsequent sections of this paper, **the types of applications that are listed above require WAN services with significantly higher bandwidth and lower latency than is currently available from wireless operators.**

5.0 Complexity of Applications

As was previously mentioned, the complexity associated with enterprise applications has a significant impact on the infrastructure that is required in order to ensure appropriate application performance. The purpose of this section of the white paper is to provide some insight into the sources of that complexity.

[1] Bridging LANs with Wireless WANs: Overcoming Technology Challenges for Wireless Data, the Yankee Group, 2002

In order to demonstrate the complexity of enterprise applications, this white paper will discuss in detail a representative enterprise application: SAP. Note that whenever this paper refers to creating a model of an application, it is actually referring to two related, but different activities. Those activities being:

1. The creation of a mathematical or analytic model that can be used to predict the performance of an application

2. The benchmarking of an application

An overview of the SAP R/3 computing architecture can be found in many references [2,3]. The computing architecture associated with SAP R/3 consists of three components. Those components are the:

1. Application Server
2. Database Server
3. Client Software, sometimes known as SAPGUI

It may seem relatively simple to model the three components listed above. **However, one of the key sources of application complexity is the sheer number of implementation alternatives that exist.**

For example, while it is common for the application and database servers to run on separate machines, in some cases these servers run on the same machine. Other implementation alternatives include the number and type of application server used (Sun, HP, etc.), the operating system (Windows, UNIX, Linux, etc.), the amount of RAM and cache in both the application server and the database server, as well as the choice of data base; i.e., Oracle, IBM DB2, Informix.

A white paper written by BMC Software, Inc.[4] developed a capacity planning model for the computing component of the SAP R/3 Sales and Distribution application. The BMC paper compared the results that were predicted by their capacity planning model to values that were measured from a test system. In general, the results that were predicted by the model corresponded very closely to the measured values.

However, as noted in the BMC paper, periodically during the test period the test system was undergoing additional maintenance that was neither related to the test nor known to the people conducting the test. The BMC paper concluded that because of this maintenance task "SAP requests were forced to wait for long periods of time resulting, among other things, in busy work processes, and generating [SAP] dispatcher wait time." As a result of the fact that a maintenance task was periodically executing, the actual response time would "jump up several seconds".

---

[2] Will, Hienger, Strassenburg, Himmer, SAP R/3 Administration, Addison-Wesley 1997
[3] Bancroft, Seip, Sprengel, Implementing SAP R/3, Manning 1996
[4] Digging into SAP R/3 for Capacity Planning, Yefim Somin. BMC Software, Inc.

The observation made in the BMC paper points out one of the other key aspects of application complexity. That aspect is that **production applications run in dynamic environments in which there is frequent, and often unpredictable, interactions between the system components.**

The preceding concept was further exemplified in a paper written by Appliant[5]. In that article, the authors described an experiment that they conducted to quantify the performance of SAP R/3. In one scenario, the authors ran a dialog in a configuration in which the only task on the client machine was running the SAPGUI program. In this scenario, the response time as seen by the user was roughly one tenth of a second. The authors then varied the scenario and had two processes running at the same time on the client machine. In this scenario, the response time as seen by the user was roughly four seconds.

A third driver of application complexity is the degree of application customization that occurs, along with the impact of this customization. In particular, most companies make changes to the standard application before the deploy it. In many cases, **even a minor change to the standard application can result in the modified system behaving significantly different than the standard application.**

6.0 A Representative Application

The preceding section identified three of the primary sources of application complexity. This section will quantify the performance of a production application and demonstrate how **a small change in WAN latency can have a major affect on the performance of an application**.

Karl Wagner, Director of Global Network and Telecommunications for PricewaterhouseCoopers' (PwC's) internal IT organization, commented that PwC often tests applications before they are implemented. For example, PwC performed extensive benchmark testing of PeopleSoft prior to deploying it. Wagner also commented that in spite of the testing that was done, PwC continues to get error messages on some PeopleSoft transactions if there is too much WAN latency.

Wagner also described an application that was internally developed by PwC. Throughout this white paper, this will be referred to as The PwC Application. Like PeopleSoft and SAP, The PwC Application is comprised of a number of differing transactions.

Table 1 contains the results of the benchmark tests that PwC performed on 3 of the transactions that are associated with The PwC Application. Note that the tests were performed against a server with moderate utilization, and that individual tests were made of each application type.

There are some interesting observations that can be drawn from the results of testing each of the three transactions.

[5] Nikolas Gloy, J. Bradley Chen, Service Level Directed Management of SAP R/3, Appliant, Inc.

Transaction #1: The Desired Maximum Response Time is 5 Seconds

Adding network delay has virtually no impact on application response time as long as the network delay is 450 ms or less. At that point, the performance of the application deteriorates badly. In particular, when the network delay is increased by 50 ms (from 450 ms to 500 ms), the overall response time goes up by 8 seconds.

Transaction #2: The Desired Maximum Response Time is 5 Seconds

Adding network delay has little impact on application response time as long as the network delay is 150 ms or less. As the network delay exceeds 150 ms, the application response time degrades rapidly.

Transaction #3: The Desired Maximum Response Time is 40 Seconds

Adding network delay has little impact on application response time as long as the network delay is 200 ms or less. As the network delay goes from 200 ms to 300 ms, the application response time exceeds the desired maximum target response time. In addition, when the network delay exceeds 300 ms, the application response time degrades very rapidly.

	Transaction #1	**Transaction #2**	**Transaction #3**
Desired Maximum Response Time (seconds)	5 seconds	5 seconds	40 seconds
Network Latency (msec)			
0	2	2	27
50	2	3	29
100	2	3	29
150	2	4	29
200	3	7	29
250	2	7	48
300	3	12	43
350	4	18	82
400	5	21	140
450	4	34	155
500	12	57	153
550	13	57	451
600	23	56	503

Results of Benchmarking the PwC Application
Table 1

It is worth noting that the section of Table 1 that corresponds to network latencies of 200 ms or more is highlighted as an avoidance zone. The phrase "avoidance zone" refers to latencies that are detrimental to the performance of the application. In particular, **since all three transactions have to run over the same wired or wireless WAN infrastructure, it is critical to provide the latency characteristic that is required by the most demanding transaction; i.e., 150 ms or less.**

7.0 ERP Application Requirements

The combination of application modeling and experience with production systems has created a number of rules of thumb about application performance. The purpose of this section is to identify some of these rules of thumb relative to Enterprise Resource Planning Applications (ERP).

7.1 SAP

The SAP HP Competence Center (The Center) has documented a number of rules of thumb relative to SAP R/3[6]. Relative to the standard OLTP user, The Center recommends 20 Kbit/s of bandwidth for each user that is utilizing WebGUI with Terminal Server, and 40 Kbit/s for each user that is utilizing WebGUI.

However, The Center provides a different rule of thumb if the user is accessing certain SAP applications, such as the Advanced Planner and Optimizer (SAP APO) application, or the SAP Business Information Warehouse[tm] (SAP BW) application. In either of those cases, The Center recommends five times as much bandwidth; i.e., 100 Kbit/s of bandwidth for each user that is utilizing WebGUI with Terminal Server, and 200 Kbit/s for each user that is utilizing WebGUI.

7.2 PeopleSoft

INS has developed a model that can be used to gain insight into the maximum size of a client-server application that can run over a given network and still provide the desired application performance. Throughout this white paper, that model will be referred to as The INS Model. A description of The INS Model is contained at www.ins.com.

In order to demonstrate the impact of bandwidth on the performance of PeopleSoft, The INS Model was run twice: once with a bandwidth parameter of 128 Kbps, and once with a bandwidth parameter of 512 Kbps. In each case, the model evaluated four PeopleSoft transactions in both 2-Tier and 3-Tier versions of the transactions.

When the model was run with the network bandwidth parameter of 128 Kbps, the maximum sized client-server transaction that could successfully take place under optimal assumptions was roughly 100 Kbytes, and only three of the eight possible transactions executed successfully. When the model was run with the network

[6] Short Introduction to mySAP Sizing, www.hp.com

bandwidth parameter of 512 Kbps, there was a notable improvement in the performance of the application. In this case, the maximum sized client-server transaction that could successfully take place under optimal assumptions was roughly 500 Kbytes, and six out the eight possible transactions executed successfully.

7.3 Amdocs

According to Amdocs, the type of network connection between the user and the ClarifyCRM server, and the latency of that network may have a major impact on performance. In order to provide the desired application performance, the following network connections and latencies are assumed

- LANs are assumed to have a bandwidth of 100 MBPS and standard latency of less than 10 milliseconds
- WANs are assumed to have a minimum bandwidth of 1.5 MBPS and latency of 150 milliseconds or less

7.4 Summary

Similar to the application described in Section 6 of this white paper, **ERP applications have an average sustainable latency requirement of 150 ms or less.** The bandwidth requirements of ERP applications range from a low of 40 Kbps to a high of 1.544 Mbps, **with the bulk of ERP applications having an average sustainable send/recieve throughput requirement in the range of 200 Kbps to 512 Kbps.**

8.0 Voice and Video

There are a variety of encoding schemes for voice (i.e., G.711, G.729) that result in different bandwidth requirements. Including the protocol overhead, it is very common for a VoIP call to require between 32 Kbps and 64 Kbps of bi-directional bandwidth.

Relative to latency, **the ITU recommends that voice should have a one-way latency of 150 ms. or less.** Other key parameters of a VoIP call are jitter and packet loss. In particular, once jitter gets above 30 ms. for a second, or packet loss gets above 2%, the quality of the voice call deteriorates quickly and significantly.

Video conferencing has differing requirements for the audio and the video components. For example, the audio component of a video-conference requires between 16 and 64 Kbps. The video component of a video-conference requires between 320 Kbps and 1 Mbps.

In summary, **the latency requirement of voice and video is similar to the latency requirements of all of the applications that have been previously discussed; i.e., an**

average sustainable latency of 150 ms or less. The sustainable send/receive throughput requirements range from a low of 32 Kbps to a high of 1 Mbps.

9.0 Email

As was mentioned in Section 4 of this white paper, email is one of the primary applications that IT organizations want to run over a wireless WAN. Relative to email systems, there are three key inter-related performance metrics. They are:

- The amount of time that it takes to send the email
- The size of the email
- The effective throughput of the WAN

Table 2 demonstrates how the amount of time it takes to send or receive a single email varies based on both the size of the email and the effective throughput of the WAN link. Note that due to factors such as packet re-transmissions, the effective throughput of a WAN link is always less than or equal to the speed of the WAN link.

Size of Email in Bytes	32 Kbps	48 Kbps	64 Kbps	128 Kbps	256 Kbps	512 Kbps
250,000	63	42	31	16	8	4
500,000	125	83	63	31	16	8
1,000,000	250	167	125	63	31	16
2,000,000	500	333	250	125	63	31
3,000,000	750	500	375	188	94	47
4,000,000	1,000	667	500	250	125	63
5,000,000	1,250	833	625	313	156	78

Email Delay in Seconds as a Function of Effective WAN Throughput
Table 2

In order to draw any conclusions from the data in Table 2 it is necessary to make some assumptions. For example, assume that an individual email must be sent or received in 60 seconds or less. The highlighted section of Table 2 indicates the situation in which that is possible. In particular, **if the email is one megabyte in size, the WAN must have an average sustainable send/receive throughput of 256 Kbps or greater. If the email is three megabytes in size, the WAN must have an average sustainable send/receive throughput of 512 Kbps or greater**.

10.0 Call to Action

As mentioned in Section 3, applications management is the most difficult component of network and systems management. The difficulty associated with applications management stems in large part from the complexity of the application itself combined with the fact that if any component of a complex application is under-provisioned, it can have a significant impact on the performance of the application.

There are some steps that an IT organization can take to improve their ability to manage applications. Those steps include:

Divide and Conquer

Instead of managing applications in a monolithic fashion, the task of applications management must be decomposed into sub-tasks.

One option is for an IT organization to decompose applications management into a network component and a compute component. Assuming that the IT organization had a target response time for the application, it could then assign a maximum amount of delay for the network, and a maximum amount of delay for the compute infrastructure. The IT organization must now ensure that it can manage its network and its compute infrastructure to meet these delay goals.

Deploy Intelligent Bandwidth

Section 3 of this paper quoted some market research performed by Network World and Packeteer that indicated that the most common technique that companies plan to use in order to ensure acceptable application performance is to add bandwidth. As was demonstrated in this paper, there are instances in which applications behave poorly because of issues with the compute infrastructure; i.e., multiple processes executing on the same server or client device. In these instances, adding bandwidth is unlikely to have any affect on the performance of the application.

However, as was also demonstrated in this white paper, many common applications require a significant amount of bandwidth and only a modest amount of latency. Yet particularly on mobile networks, there will never be enough bandwidth and it will never be inexpensive enough to just mindlessly throw bandwidth at an applications problem. The solution is to deploy WAN services that have appropriate throughput and latency characteristics, as well as QoS functionality.

When deploying a wired or wireless WAN, it is necessary to support the latency requirement of the most demanding application or transaction. As was demonstrated in this white paper**, many common applications require a sustained average latency of 150 ms or less in order to function properly.**

When deploying a wired or wireless WAN, it is also necessary to support the aggregate throughput requirement of all of the applications that will run simultaneously over the WAN. Considering the mix of applications that IT organizations want to run over a wireless WAN together with the bandwidth requirements of these applications, **a successful wireless WAN service must provide an average sustainable send/receive throughput of at least 300 Kbps.**

Dr. Jim Metzler

Dr. Jim Metzler is widely recognized as an authority on both Information Technology (IT) and its business applications. In over 28 years of professional experience, Jim has assisted numerous vendors to refine their product strategies, multiple service providers to deploy technology and services, and simultaneously helped tens of enterprises evolve their IT infrastructure.

Jim's current service provider interests include the rapid deployment of profitable services; Business Case Analysis, and the implementation of effective Operation Support Systems. His current enterprise interests include the evaluation and deployment of new technologies (Voice over IP and IP-Based VPNs), infrastructure optimization, the use of outsourcing, as well as proactive network and systems management.

Jim has performed market research into network and systems management issues for a number of vendors, including Cisco, Nortel, Visual Networks, and NetScout. He recently helped a Fortune 200 company evaluate outsourcing the management of their US based network. Jim has helped multiple companies to develop processes to baseline their IT infrastructure and has also assisted companies to evaluate the deployment of usage sensitive chargeback.

In November of 2003, Jim will be the keynote speaker on an IT management seminar tour produced by NetScout. In February 2004, Jim will be the keynote speaker on an IT management seminar tour produced by Network World.

Jim just completed an assignment in which he helped a network equipment vendor to position a new product in the marketplace. He is in the final stages of a project to help a professional services firm evaluate the viability of offering managed security services. On multiple occasions, Jim has served as an expert witness, most recently in a case between a Fortune 50 company and the company to whom it had outsourced its IT infrastructure.

In the recent past, Jim has published four articles in Business Communications Review (BCR) on the topics of Security, Service Level Management, Wide Area Networking, and VoIP. Jim also recently published two articles in Network World on the topic of Metropolitan Area Networks, and will soon be publishing an article in Network World on the topic of Storage Area Networks.

Dr. Metzler has worked in many positions in the IT industry. This includes being a compiler writer for a branch of the US intelligence community; creating software tools to design customer networks for a major IXC; being an Engineering Manager for high speed data services for a major Telco; being a Product Manager for network hardware; managing networks at two Fortune 500 companies; directing and performing market research at a major industry analyst firm; and running a consulting organization.

Jim holds a Ph.D. in Numerical Analysis from Boston University. He is a member of the Networld + Interop Planning Committee. He has co-authored a book, published by Prentice

Hall, entitled “Layer 3 Switching: A Guide for IT Professionals” which was recently translated into Chinese.

Dr. Metzler has been on the faculty of a number of universities, including Boston University, Northeastern University, Bentley College, and Drew University. At the Networld + Interop conference in both Las Vegas and Atlanta Jim runs a two day conference entitled “Strategic Interop”. The goal of this two-day conference is to explore how Information Technology can be used to meet business objectives.