Why SAP Performance Needs Help

By Peter Sevcik and Rebecca Wetzel
July 2006

Consider this. An employee entering data into an SAP system proceeds quickly and efficiently when the system responds within two seconds. As response times exceed four seconds the employee begins to roll his eyes and mutter under his breath, and when response times stretch to eight seconds, the employee becomes frustrated and takes an unscheduled coffee break. Business productivity suffers as a result.

This scenario plays itself out every day across the globe, and sums up why it is important that SAP applications perform well over a wide area network (WAN). Unfortunately, forces such as network congestion, distance, and denial-of-service attacks conspire to degrade SAP-over-WAN performance to frustrating levels. It is, therefore, vital to protect and improve SAP performance using a variety of performance optimization tools such as those providing Quality-of-Service (QoS), data compression, and server offloading capabilities.

Business Importance of SAP Performance

Unlike most software applications, SAP is business critical by nature. While other applications like mySQL and Oracle/Peoplesoft support a gamut of uses ranging from ‘nice to have’ to ‘mission critical’, SAP applications are always used to run the business—the lifeblood of the company.

SAP applications are as ubiquitous as they are mission critical. With over 33 thousand customers in more than 26 industries around the world, SAP is the world’s largest business software company, and the third largest independent software provider.

NetForecast’s research has determined that a response time of two seconds is a common requirement to ensure SAP user productivity, and user tolerance for long response times is low. If SAP performance suffers, so does the business, therefore, performance expectations and requirements for SAP are more stringent than for other business applications.

As the lifeblood of the company, there is no such thing as SAP “lite”. SAP is always in the business limelight, so it simply must perform well over the WAN.

SAP Performance Characteristics

SAP applications support two graphical user interface (GUI) versions, each with unique performance characteristics. The first, which we refer to here as SAP-GUI (a.k.a. SAP R/3), has a SAP-proprietary user interface—and the second, which we refer to as SAP-Web (a.k.a. mySAP), has a Web-enabled user interface. Most businesses use SAP-GUI today, however, the popularity of the Web-enabled version is growing, and many SAP customers are migrating from the SAP-GUI to the SAP-Web user interface.

Customers generally characterize SAP as a ‘good network citizen’ compared to other applications. To place SAP performance into perspective, it is useful to compare performance profiles for SAP tasks with other common applications, some of which are not exemplary network citizens.

Figure 1 shows how various applications differ in the number of application turns (round trip client/server protocol interactions) per task and payload per task on a logarithmic scale. Note that NetForecast created the profiles in Figure 1 using actual, not theoretical profile data for typical user tasks.

NetForecast Report 5084

The average payload remains nearly identical for both SAP versions, but the turns per task do not. Turn counts for the SAP-Web version are higher than for its SAP-GUI counterpart (as often happens when applications are Web enabled). In fact, SAP-GUI averages a mere four turns per task, whereas SAP-Web averages more than twice that at 10 turns per task. The higher SAP-Web turn count exacts a performance toll for users, in large measure because each additional turn adds delay to the performance equation, especially as distance across the WAN increases.

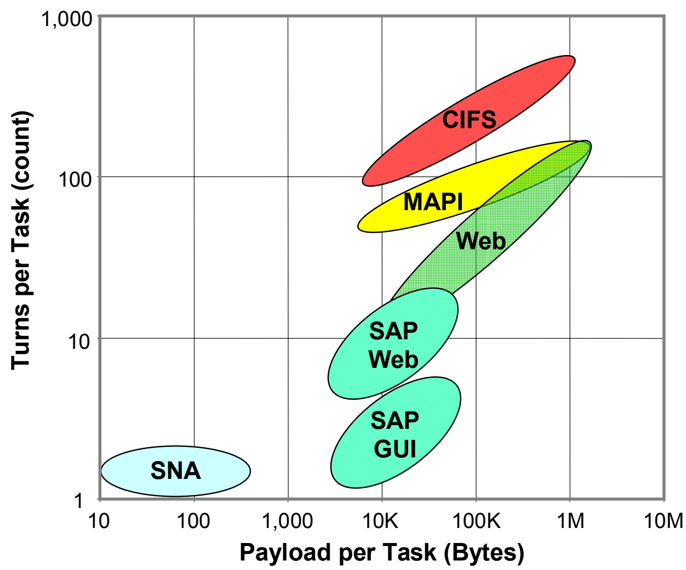

Figure 1 – Application Profiles

Average turn counts and payload sizes for both SAP versions are higher than for SNA, which wins the prize for best network citizen because it is streamlined and lean. Compared to Microsoft's Common Internet File System (CIFS) and Messaging Applications Programming Interface (MAPI)[1], however, SAP performs quite well. SAP also compares favorably to most business-to-business Web pages.

SAP-over-WAN Performance Threats

Although SAP is a well-behaved network citizen compared to most applications, it cannot escape a number of common WAN performance challenges and security threats. Chief among these are congestion, distance, and denial-of-service attacks. Although these challenges and threats affect all applications equally, they assume increased importance because SAP is business critical, and carries more rigorous performance requirements than other applications.

Understanding Application Response Time

To understand how congestion, distance and denial-of-service attacks impair application performance, it is important to understand key factors influencing application response time. These include: application turns, application payload, network bandwidth, network round-trip time (RTT), server compute time (Cs), and client compute time (Cc).

[1] *For information about CIFS and MAPI performance characteristics and performance solutions see NetForecast report NFR5081 "Why Centralizing Microsoft Servers Hurts Performance".*

The following formula shows how these factors combine to determine a user's response time for a given task (see the *Glossary of Terms* at the end of this report for detailed formula definitions).

$$ResponseTime \approx Turns(RTT) + \frac{Payload}{Bandwidth} + Cs + Cc$$

Congestion as a Threat

SAP is only one among a burgeoning throng of applications contending for WAN bandwidth. Barring special treatment, SAP traffic must compete on an equal footing with all other applications in a WAN-resource free-for-all. SAP must also contend with congestion from such traffic sources as virus and worm propagation.

Without protection, SAP can readily be starved of essential bandwidth, and as the above response time formula shows, when SAP bandwidth is squeezed, user response times rise—as do those unscheduled coffee breaks.

Distance as a Threat

SAP is generally centralized rather than distributed, forcing remote users to access the application over long distance WANs. Long user-to-server distances bump up round-trip times and slow application response times for users.

If you do not live in the world of SAP, here are some reasons why SAP centralization is the norm. Multiple SAP instances increase the likelihood of data discrepancies, such as the possibility that a product part number, price, or other critical information may be changed in one database but not others. Also, SAP software is expensive and requires specialized skills to manage it. When you centralize SAP deployments, fewer SAP experts can support more users, and you can save money on the software as well.

In addition, SAP centralization makes it easier to document, track, and control information to comply with regulations like the US Sarbanes-Oxley Act or Basel II.

Denial-of-service Attacks as a Threat

As enterprises migrate to SAP-Web, users are increasingly vulnerable to the effects of denial-of-service attacks directed at Web servers. Such attacks can tie up server resources and deny or degrade service to legitimate users. This problem is especially acute when employees or business partners must access SAP-Web servers over the public Internet.

How Apdex Works

Apdex provides a numerical measure of user satisfaction with the performance of enterprise applications. Apdex reports are arrived at using a four-step method.

Step 1 – Select Target Time

Define a target response time of T seconds. Response times between 0 and T are considered satisfactory by the users within a business scenario.

Step 2 – Measure Performance

Measure performance and place each measurement into one of three user-perceived performance zones: satisfied (Sat), tolerating (Tol), or frustrated. The satisfied zone is defined in Step 1. Frustrated times are above F seconds, where F=4T. Tolerating times are between T and F seconds.

Step 3 – Apply the Apdex Formula

Sum the incidents of response times that fall into each zone, and divide by the total number of measurements using the formula (the sum of frustrated incidents is not in the numerator but is included in the denominator):

$$Apdex = \frac{Sat + Tol/2}{Total}$$

Step 4 – Show Results

Show the Apdex Index value (on a scale of 1 to 0) together with the corresponding threshold T. The Apdex Scale is subdivided into performance quality regions ranging from excellent to unacceptable.

In this report, we describe how T is selected for each application (Step 1). For brevity, the details of Steps 2 and 3 are not shown. Results of Step 4 illustrating the Apdex Quality Regions are shown in Figure 2.

Apdex is an open standard developed by the Apdex Alliance. Visit www.apdex.org to learn more.

NetForecast Performance Model

NetForecast has studied application performance for over a decade and has embodied that knowledge in a detailed model accounting for the performance factors that affect user response times. The NetForecast model uses a multi-variant Monte Carlo analysis technique that applies real-world distributions to parameters in the response time formula.

Among the factors accounted for in the model is the inescapable fact that round-trip time is directly proportional to distance divided by the speed of light—therefore, the longer the distance, the longer the round-trip time. The model also takes into account the fact that packet loss escalates with distance because the risk of congestion increases with higher network hop counts. The model can also show effects of bandwidth contention among applications, as well as the consequences of other aspects of network stress such as high network latency and loss.

NetForecast has found that it is unhelpful to compare raw response time values for each application because task definitions and judgments about acceptable task completion times vary widely for each application type. For example, what might be a painfully slow task time for one application task type might be entirely acceptable or even excellent for another.

Therefore, to equalize the view of performance across applications (in this case SAP-GUI and SAP-Web) and network scenarios, we apply an application performance index methodology called Apdex, developed by the Apdex Alliance as an industry standard for measuring and assessing user satisfaction with application performance (see the *How Apdex Works* sidebar on page 3).

SAP application tasks are highly repetitive and designed for short query responses, and SAP users typically require a target time of two seconds to remain productive. So we assigned SAP-GUI and SAP-Web an Apdex target of two seconds (Apdex T=2). This contrasts to typical target response times of one second for SNA, six seconds for business-to-business Web sites, nine seconds for Microsoft MAPI, and 12 seconds for Microsoft CIFS.

The model applies a distribution of server response times centered on a one-second average. This leaves another second for the network to deliver the complete task within the two second target. NetForecast has seen SAP applications that achieve one second at the data center as modeled here. But we have also seen SAP implementations that must dig through extremely large databases for many of the task responses and deliver the task reply in six seconds or more. In such cases, the Apdex target T must be scaled up appropriately.

The relative performance and conclusions drawn in this report would not change by applying a larger T value appropriate to longer server response times.

Baseline Performance

The NetForecast model in Figure 2 shows the effects of server-to-user distance on SAP-GUI and SAP-Web performance under optimal terrestrial network conditions on a high-quality private WAN.

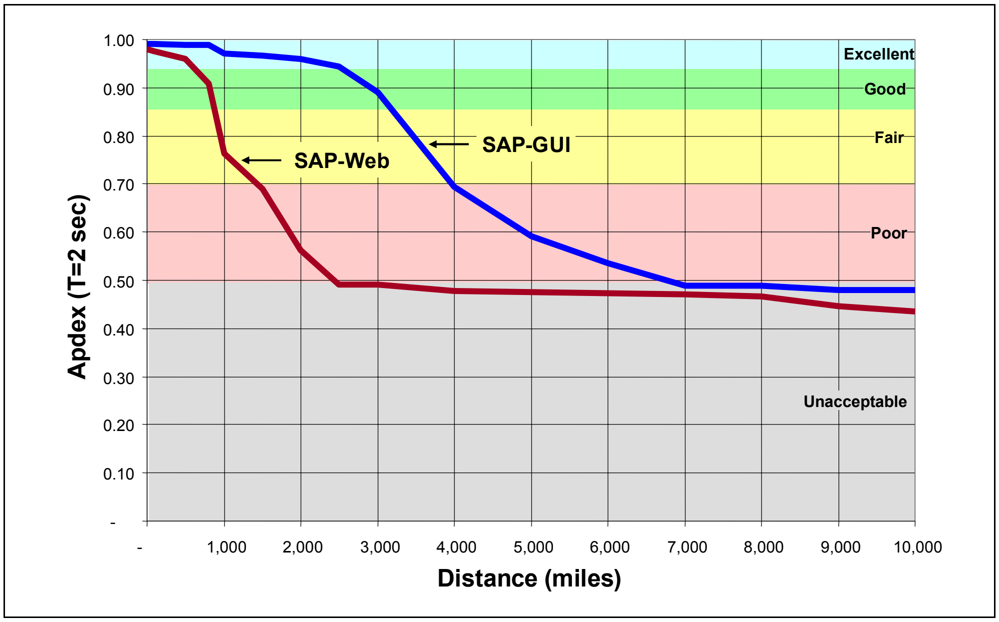

Figure 2 – SAP Performance Baseline Curves

These baseline results show that SAP-Web pays a performance price for more than doubling the number of turns per task compared to SAP-GUI. Because of the higher turn count, SAP-Web performance deteriorates more quickly with distance than its SAP-GUI counterpart. SAP-Web performance dives into the poor range within only 1,500 miles of the server, whereas SAP-GUI performance slips into the "poor" range within 4,000 miles of the server.

SAP-Web is routinely deployed using Secure Sockets Layer (SSL) encrypted links over the public Internet. Baseline performance over the Internet will be much poorer than the optimal WAN baseline shown here.

How to Protect and Improve SAP Application Performance

SAP-GUI and SAP-Web applications can be protected from performance problems caused by congestion and other network stresses that force the baseline performance curves shown in Figure 2 dramatically down. A downward-shifted curve is bad news for users because the lower the Apdex score, the poorer the user experience. SAP-Web can also be protected from denial-of-service attacks that illegitimately compete for server resources.

Implementing performance protection assures an optimal user experience, even in the face of challenging network conditions.

In addition to protecting the user from performance degradation, the SAP user experience can be improved by applying acceleration techniques that shift the baseline performance curves up. Shifting performance curves up improves the experience for far-flung users.

Two application acceleration approaches can be brought to bear on SAP performance problems. A distributed approach (a.k.a. dual-ended or symmetrical) brings enormous benefits to SAP-GUI users, and a centralized approach (a.k.a. single-ended or asymmetrical) specifically benefits SAP-Web users.

We refer here to solutions using the distributed approach as WAN Optimization solutions, and those using the centralized approach as Application Front End (AFE) solutions. WAN Optimization solutions require products at both ends of a WAN connection, while AFE solutions require a product in front of the server where the application resides, and a Web browser at the user's end of the WAN connection.

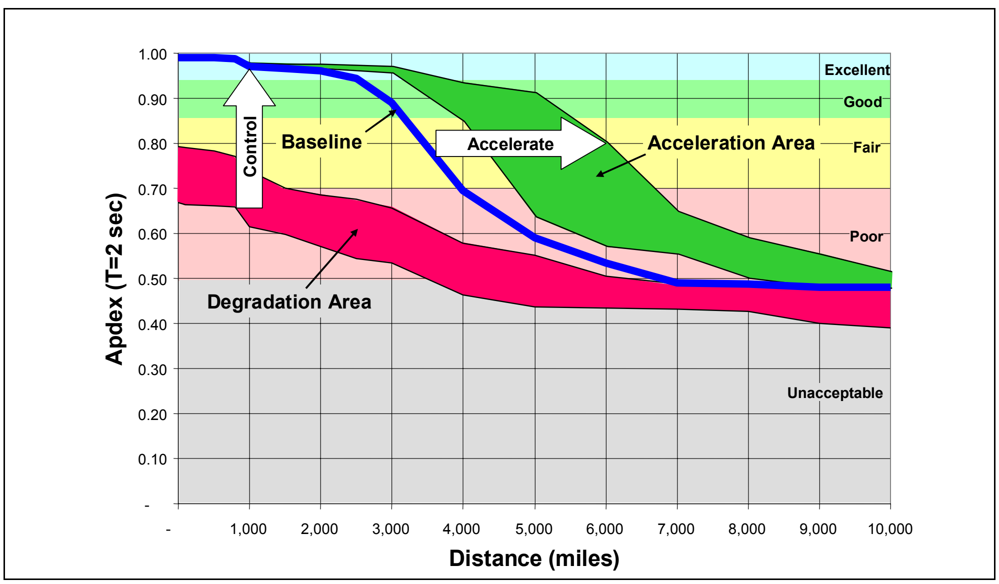

Figure 3 – Effects of Dual-Ended WAN Optimization of SAP-GUI

Because it is difficult to manage application performance without visibility into that performance, both centralized and distributed approaches must build on a foundation of traffic monitoring and application performance measurement. Visibility enables solutions to apply application control techniques that restore degraded performance to the original baseline, and to apply acceleration techniques to improve performance beyond the original baseline, thus moving the performance curve up and to the right.

WAN Optimization Solutions - Ideal for SAP-GUI Optimization and Acceleration

WAN Optimization solutions are ideally suited for making and keeping WAN-connected SAP-GUI users fully productive—even thousands of miles from the application server. Distributed ADS solutions ***optimize*** performance by applying Quality-of-Service (QoS) techniques such as traffic prioritization and bandwidth shaping to SAP-GUI traffic traversing the WAN to prevent performance curves from slipping into the "degradation area" shown in Figure 3. WAN Optimization solutions also can ***accelerate*** performance through compression and TCP acceleration to drive performance curves into the "acceleration area". Figure 3 shows the results of a distributed solution applied to SAP-GUI traffic.

It is useful to note that QoS techniques help assure optimal performance for all users regardless of how far they are from the server, whereas acceleration techniques improve response times primarily for remote users.

The benefits your users experience from distributed solutions will vary depending on your network type and conditions, as well as the specific product(s) you choose. For example, acceleration techniques applied to a satellite network are likely to yield more dramatic performance improvements than the results for a terrestrial network shown in Figure 3.

We would like to add a word of caution here about using acceleration techniques without also deploying QoS for control. Applying acceleration in the absence of QoS is not recommended because despite the benefits of acceleration, SAP can easily slide into the degradation area under adverse network conditions even if SAP traffic is compressed and accelerated.

AFE Solutions Protect and Improve SAP-Web

An AFE solution in the data center can overcome SAP-Web's turn count disadvantage by reducing the number of round trips per task, and thus improving the user's response time into the acceleration area shown in Figure 4. HTML transformation in the AFE can lower the turn count for each SAP task. By lowering the turn count, the centralized approach nudges SAP-Web's WAN performance curve to within kissing distance of the SAP-GUI curve.

An AFE solution is an absolute necessity for Internet-connected users who will experience a much poorer baseline performance curve than that shown here. Without an AFE solution, SAP-Web users will be subjected to the variances of Internet performance with nothing to ensure their productivity.

An AFE solution also can prevent performance from slipping into the degradation area by acting as a buffer between outside clients and back-end SAP servers, terminating all incoming and outgoing sessions. This prevents outside sources from identifying or accessing the application servers and obtaining knowledge of the transactions between the client and servers. The solution also needs to include application firewall features to fight distributed denial-of-service attacks.

Because it is single sided and does not require a companion device on the user's end, AFE solutions can be used over the public Internet as well as over a private WAN.

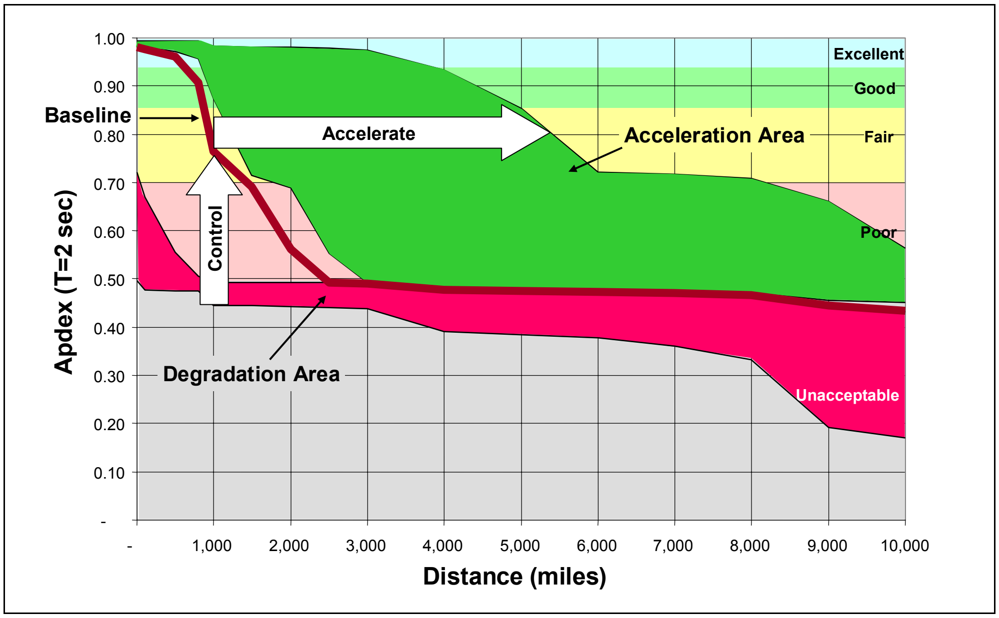

Figure 4 – Effects of Single-Ended WAN Optimization of SAP-Web

Combining Solutions Ensures Consistency for All WAN Users

Figure 5 shows how WAN Optimization solutions applied to SAP-GUI, combined with AFE solutions for SAP-Web, equalize top-line performance for both SAP interfaces. This means that users will enjoy a consistent WAN experience regardless of which graphical user interface they use.

A consistent user experience is especially important for enterprises that are migrating WAN-connected users from SAP-GUI to SAP-Web. Without solutions in place to make the experience comparable, SAP users accustomed to excellent performance of SAP-GUI are likely to be frustrated when they transition to non-optimized SAP-Web. Therefore, NetForecast recommends that enterprises employ WAN Optimization solutions for both SAP flavors, plus AFE solutions for SAP-Web.

Applying application acceleration products to both the GUI and Web versions of SAP has a dramatic effect in top-line performance. SAP is transformed from providing good performance within a continent and unacceptable performance across the globe, to good performance across multiple continents and acceptable performance anywhere. In our experience, once an application acceleration solution is adopted, it is never removed.

Not only will application acceleration help SAP, it will also bring similarly positive performance improvement to other business-critical applications.

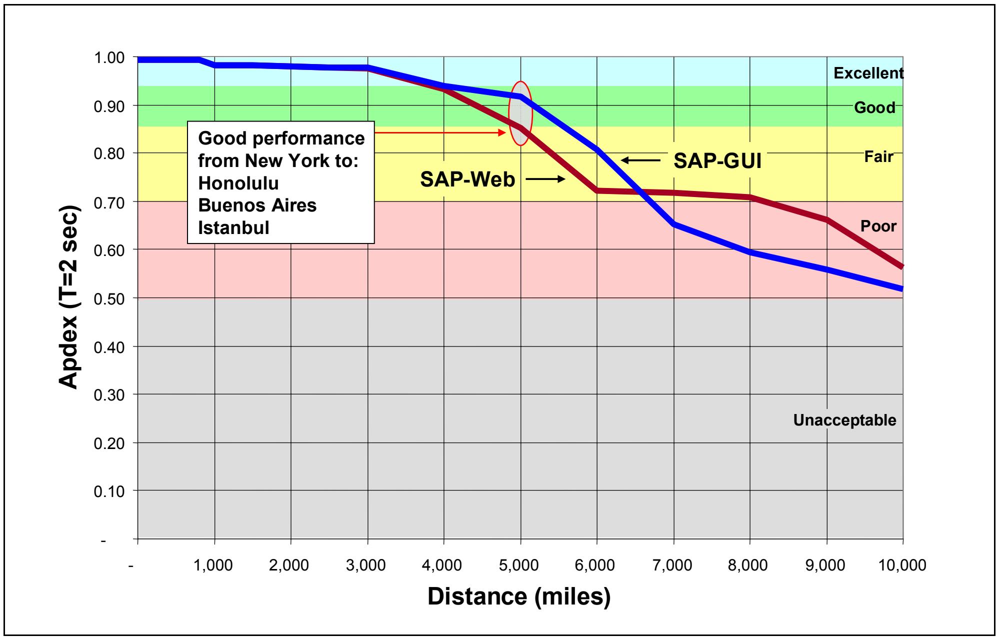

Figure 5 – Maintaining Consistent High Performance for SAP-GUI and SAP-Web

Conclusion

To ensure that your SAP users are fully productive and are not plagued by performance-related frustrations, NetForecast recommends implementing solutions that optimize the performance of your SAP-GUI and SAP-Web applications across the WAN. Fortunately, there is much you can do to protect and improve SAP-over-WAN performance—and because SAP is integral to business, it makes sense to do so.

If you are using the SAP-proprietary graphical user interface, consider deploying a WAN Optimization solution with Quality-of-Service capabilities, as well as application acceleration.

If you are using the Web-enabled version of SAP over a WAN, we also recommend a WAN Optimization solution, and we suggest that you combine it with an AFE solution that provides additional acceleration and server offload capabilities—and if you are using SAP-Web over the Internet, an AFE will be an absolute necessity to ensure SAP user productivity.

Finally, to ensure that your ADS system is optimizing and accelerating SAP applications as effectively as possible, NetForecast recommends that you deploy tools to continuously monitor and report on SAP performance. Monitoring and reporting will allow you to assess SAP performance results, ensure that Quality-of-Service parameters and other system settings are appropriately configured, and will give you information that can be used to troubleshoot and improve the system. Such actions are especially important to assure the productivity of distant users.

As the adage goes, your mileage may vary from what our results show. But we can say with confidence that your mileage will improve when you apply a dual and/or single-ended application delivery system solution to keep your SAP users happy and business productivity high.

Apdex:
Apdex is an open performance reporting standard defined by the Apdex Alliance. See: www.apdex.org

Glossary of Terms

Bandwidth is the minimal bandwidth (bits per second) across all the network links between the user and the application server. The slowest link is typically the user's access line to the network. Useable bandwidth of the link may be reduced by the effects of conflicting traffic (congestion) and protocol efficiency (e.g., TCP window).

Cc (Compute Client) is the total processing time (seconds) required by the client device.

Cs (Compute Server) is the total processing time (seconds) required by the server(s).

F The threshold which defines the boundary between tolerating and frustrated performance zones.

Frustrated Zone Any application response time above F in which the user is very negatively affected by response time.

Payload is information content (bytes) that must be delivered to/from the user's device.

Performance Zone A range of time (between two time values) that a user waits for an application to respond during which his/her perception of the application's responsiveness does not change.

Response Time is defined as the elapsed time (seconds) between a user action (e.g. mouse click, enter, return) and the system response (client, network, server), so the user can proceed with the process. The aggregation of these individual task completion waiting periods defines the "responsiveness" of the application to the user.

RTT is the round-trip-time (seconds) between the user and the application server.

Satisfied Zone The range of application response times between 0 and T in which the user is not affected by the response time.

T The target threshold which defines the boundary between satisfied and tolerating performance zones.

Task Each user-application interface interaction that requires a user entry and an application response.

Tolerating Zone The range of application response times between T and F in which the user is negatively affected by response time.

Turns are the application client-server software interactions (turn count) needed to generate a user-level system response or task (see above). Turns are separate form and in addition to TCP two-way interactions (e.g., open, close, ACKs). The user is not aware of turns.

Peter Sevcik is President of NetForecast and is a leading authority on Internet traffic, performance, and technology. Peter has contributed to the design of more than 100 networks, including the Internet, and holds the patent on application response-time prediction. He can be reached at peter@netforecast.com.

Rebecca Wetzel is an Associate of NetForecast and a 20-year veteran of the data networking industry with unparalleled inside knowledge of the Internet service and product markets. She works with network product vendors and service providers to develop and implement product strategies. She can be reached at rebecca@netforecast.com.

NetForecast helps enterprises and vendors understand and improve the performance of networked applications.

Additional information is available at: www.netforecast.com