NetQoS
Performance Experts

WHITEPAPER

Best Practices for Monitoring Business Transactions:

Business transaction monitoring has never been more critical to operational efficiency, yet there remains much confusion over methodology. The implementation choices consist of different deployment strategies (client-site or server-site, agent, or appliance) and distinct monitoring technologies (active or passive). The different options have their individual strengths and weaknesses. This article discusses industry best practices for effectively monitoring business transactions in a global environment.

Best Practices for Monitoring Business Transactions

Business transaction monitoring has never been more critical to operational efficiency, yet there remains much marketing confusion over methodology. The implementation choices consist of different deployment strategies (client-site or server-site, agent or appliance) and distinct monitoring technologies (active or passive). The different options have their individual strengths and weaknesses. This article discusses industry best practices for effectively monitoring business transactions in a global environment.

Deployment Strategies

One of the most important decisions is the deployment strategy for the business monitoring solution. Should monitors be deployed at the client sites or should they be deployed at the data centers? Should software agents or hardware appliances be used? While this may seem like a minor matter, it has serious ramifications from an immediate headache and recurring cost standpoint.

Client-site Approaches

Client-site approaches require that software be installed on clients' desktops or hardware be installed at the clients' sites. For large enterprises, this approach may prove to be a deployment and management headache that requires cooperation among multiple management fiefdoms. The individual who manages the network is often different from the individuals who manage the desktops at the various sites. Indeed, deploying either software or hardware at client sites might not be possible and is rarely easy. It can also be quite painful to maintain large numbers of remote monitors, keeping them continuously running and up-to-date.

From a technical standpoint, client-site approaches have some important weaknesses. First, they usually have a very limited view of the client-application-network environment. Because of cost, maintenance, and load issues, they are typically deployed in limited quantities across the network. They therefore only get sample, hopefully representative, measurements of the overall environment. Client-site software agents are particularly ineffective in that a single computer at a site might be selected to represent behavior for the entire site…or even multiple sites. Client-site hardware appliances (using passive monitoring technology) might be placed at the access router to measure performance for all clients at that site.

Client-site approaches may unduly stress the network or servers. The network might be stressed as remote monitors upload their performance statistics to a centralized data store. It is wise to ask the vendor for bandwidth usage metrics per monitor as a function of number of transactions—and then perform the measurements yourself to verify. If active monitoring technology is used, additional traffic is inserted from the synthetic transactions; this may unduly load the network links or the servers themselves. Unacceptable network or server stress is another reason that the number of client-site monitors is often reduced to a representative sample, resulting in a limited view of the environment.

Because of their location, client-site approaches have difficulty separating the server delay from the network delay. A common technique is to measure the network delay based on the initial TCP connection setup time and then assume that the network delay is constant throughout the session. This approach can be grossly inaccurate, particularly when persistent sessions (now common with web) or long sessions (common with telnet, FTP, and so on) are involved. It also completely ignores the effect of serialization delay because the connection setup involves the smallest sized packets. It also ignores self-induced queuing delay. Some augment their network delays by periodically actively sending ping (ICMP) packets but this approach suffers from similar drawbacks.

Server-site Approaches

Server-site approaches allow monitors to be placed at the datacenters rather than at the clients' sites. This reduces the number of monitors, greatly easing deployment and management issues. Not only are there fewer systems to manage, but datacenters will have people more experienced in their maintenance.

Special care must be taken if the approach requires that software be installed on the actual production servers. Systems managers are rightfully nervous of potential software conflicts, and some have had negative experiences with the monitoring software crashing their systems. They do not deem such as a career-enhancing event. Usually the easiest solution, and certainly the one with least risk, is one that allows a hardware appliance to be placed near the servers off of a tap or span port—rather than one that requires that software be installed on the servers themselves.

Server-site passive approaches provide a wonderful vantage point. Server-site monitors can see all users interacting with all servers at the datacenter, on a 7x24 basis, because that is where they are located. Server-site active approaches have a horrible vantage point if network information is important (and it generally is)—they only see the datacenter LAN.

Server-site monitors place much less stress on the network and servers. The performance statistics are uploaded over well-provisioned links because the (passive or active) monitors are already located at the datacenters. Likewise, the additional traffic from active monitors' synthetic transactions occurs over higher-capacity links, and the stress to servers is generally much lower because fewer monitors are needed (as compared to client-site deployment).

Because of their location, server-site approaches have no difficulty separating the server delay from the network delay. However, they will have trouble identifying client processing time from client silence. That is, they will not know whether the client CPU is busy or if the client is simply drinking coffee and chatting.

Deployment Summary

The preferred deployment strategy uses the server-site approach. It greatly reduces deployment and maintenance headaches, places minimal stress on the network, and can provide a virtually unlimited view of the environment. To reduce risk, deploy a hardware appliance to avoid installing software on the production servers. If you deploy client-site monitors, you can reduce their numbers by also deploying a server-site monitor.

Monitoring Technologies

Another important—and frequently contentious—decision for selecting a business transaction monitoring solution is whether to use active or passive technology. An active monitor emulates a client by periodically generating synthetic transactions according to some user-defined script. In contrast, a passive monitor measures the transactions of real clients in all their variability. Which is better? Well, it all depends on what you want to measure.

Active Monitors

Active monitoring approaches generate synthetic transactions. They use a form of robot to periodically perform one or more defined business transactions. The robots follow a script, a sequence of timed commands, in their interactions with the server. They are often installed on dedicated systems to minimize the number of system variables between script runs. By always running the exact same transaction in the same manner on the same platform with no other application competing for resources, active monitors provide a deterministic baseline that reflects variations in server and network performance. The client variability has been effectively removed.

The advantages of an active monitor are that you know what it is doing, and you know when it should be doing it. You know that any significant deviations in performance measurements are likely due to changes in network or server behavior. You have controlled 7x24 activity, which is useful for availability monitoring.

The disadvantages to active monitoring are that you do not know what the real users are experiencing, and the monitors can significantly degrade the real user performance. Addressing the second point first, active monitors place additional load on the network and the servers. Without careful planning, active monitors have been known to congest network links and bring servers to their knees—an avoidable but all too common situation.

While active monitors are useful for availability monitoring, load concerns usually limit their effectiveness. They are typically programmed to perform their transactions only every 15 minutes or so to prevent stressing the environment. This means, on average, they detect a failure after 7.5 minutes have passed—much better than never, but the helpdesk phone has probably already been ringing if it happens during normal business hours.

The primary disadvantage to active monitors is that they do not capture the real user experience. The script they follow might bear little resemblance to how actual clients are using the application; it is simply a model of a possible transaction. This is useful for monitoring changes in performance of the scripted transaction, but it does not necessarily relate to the real user.

Even if a user were to perform the same transaction with the same programmed timing as the active monitor, the performance of the real user might differ significantly from the active monitor because of differences in underlying software. For example, some commercial active monitors do not use a web browser when sending commands to the Web server; instead, they use an API to send requests serially within a single session. The real user will be using Internet Explorer or Firefox or the like, and will send requests simultaneously in multiple parallel sessions. The time it takes for a typical Web page to download will thus differ significantly for the real user and the active monitor—even though the Web page is the same. Differences in hardware, operating systems, drivers, and other software can impact experienced performance.

The performance reported by active agents might also differ substantially from real users due to caching or other acceleration techniques. By periodically repeating identical requests, active monitors might experience a significant performance boost from caching technologies—on the servers, on network devices, or on the client itself. While caching on the client might be disabled, caching on the servers or other network devices cannot be disabled without harming the real users. One approach that eliminates this caching benefit is to program the active monitor to send random queries, but this also destroys their deterministic advantage—you no longer know what is being measured.

Passive Monitors

Passive monitoring approaches measure real user traffic and behavior. They accommodate variations in user behavior, systems, Web browsers, and networks—they do not assume that a single model is representative. They can provide an unlimited view of performance in terms of different transactions, different network segments, different servers and different application tiers. Passive monitors can either report on individual transactions (verb monitors) or on an aggregation (generic monitors).

Verb monitors provide individual performance statistics for each configured verb, where a verb can be a URL for web applications, a specific query for database applications, or a document download for FTP applications. This approach provides the most granular performance detail at a cost of scalability and ease-of-use (because each verb must be configured). If many verbs are configured, important patterns might be hidden by the noise—the trees might obscure the nature of the forest.

Generic application monitors typically summarize the performance results for the different observed verbs. This approach reduces configuration requirements and improves scalability at a cost of reduced granularity. For example the performance of all FTP document downloads from a particular server within a specific size range (such as between 23 and 46 KBs) might be presented as a single average metric using the generic approach, whereas an FTP-specific monitor might require that you configure each document that you wish to monitor to be able to report their individual performance statistics.

The preferred passive monitoring solution combines the ease-of-use and scalability of generic monitoring with the flexibility and detail of verb (transaction) monitoring. That is, it provides out-of-the-box generic monitoring but also supports user configuration of custom verbs (transactions). Support for custom transaction configuration in passive monitors is less flexible than that in active monitors.

There are two main disadvantages to passive monitors. One is that they might not provide the flexibility to define a desired custom business transaction—the applications and transactions that they monitor might be limited. Another disadvantage is also their strength: their measurements include the variability inherent in real user behavior.

Technology Summary

Active monitors eliminate uncertainty about what is measured and provide a check for loss of service. They provide a very limited view of performance, limited by number of transactions, locations, and environments. They can receive an artificial performance boost from caching, and they do not capture the real user experience.

Passive monitors capture the real user behavior and provide a potentially unlimited view of application, network, and server performance. They lack the determinism and control intrinsic to active monitors. They might be limited in their support of custom-defined transactions.

Best Practices

The preferred deployment strategy uses passive server-site monitoring in the form of an appliance. The server-site approach greatly reduces deployment and maintenance headaches, places minimal stress on the network, and can provide a virtually unlimited view of the environment. Use of an appliance reduces risk by avoiding the need to install software on production servers.

The optimal technology approach combines both passive and active monitors. The passive server-site monitor effectively captures the real user behavior and provides a potentially unlimited view of application, network, and server performance. There is no need to deploy active monitors across the network—the passive server-site monitor will report any network performance problem, including loss of network availability. Therefore only a single active monitor placed at the datacenter is necessary to provide a deterministic baseline of custom configured transactions.

About the Author

Dr. Cathy Fulton

Co-Founder, Executive Vice President
and Chief Technology Officer of NetQoS

Dr. Cathy Fulton is a highly recognized and credentialed network performance expert. Dr. Fulton has earned many outstanding distinctions over the years, such as National Science Foundation Fellow, National Defense Science and Engineering Fellow, Navy Achievement Medal recipient, National Defense Service Medal recipient and Navy Master Training Specialist. In addition to these distinctions, Dr. Fulton is a member of the board of NetQoS, Inc. and is an active member of both IEEE (Institute of Electrical and Electronics Engineers) and ACM (Association for Computing Machinery).

Dr. Fulton is widely published in networking journals such as the IEEE/ACM Transactions on Networking and IEEE Journal of Selected Areas in Communications. She is also sought after to present her research and work for international conferences and events such as Interop, IEEE Infocom, the ATM Forum Traffic Management Group, and the Computer Measurement Group.

Prior to co-founding NetQoS, Dr. Fulton was employed as a Vice President for Smaqcom, a network consulting firm and as a Research Scientist and Project Leader for Schlumberger. Dr. Fulton has provided consulting to major firms such as Cisco, Schlumberger, and Samsung, and has taught nuclear physics at the U.S. Navy's Nuclear Power School. Since January 1999, she has been developing the NetQoS proprietary system for network performance management.

Dr. Fulton received her Ph.D. in Electrical and Computer Engineering from the University of Texas, specializing in performance analysis of multimedia networks. She also holds a B.S. in Physics from Texas A&M University.

About NetQoS

NetQoS is the fastest growing network performance management products and services provider. NetQoS has enabled hundreds of the world's largest organizations to take a Performance First approach to network management—the new vanguard in ensuring optimal application delivery across the WAN. By focusing on the performance of key applications running over the network and identifying where there is opportunity for improvement, IT organizations can make more informed infrastructure investments and resolve problems that impact the business. Today, NetQoS is the only vendor that can provide global visibility for the world's largest enterprises into all key metrics necessary to take a Performance First management approach. More information is available at **www.netqos.com**.

NetQoS, Inc.
e. info@netqos.com
p. 512.407.9443
t. 877.835.9575
f. 512.407.8629
www.netqos.com

WP rev1 20061030