Managing User Website Experience: Comparing Synthetic and Real Monitoring of Website Errors

By John Bartlett and Peter Sevcik

January 2006

NetForecast Report 5080

The modern enterprise relies on its web sites to provide information and services to its customers, partners and suppliers. As dependence on this medium increases, so increases the need to have this interaction with the company be a positive experience. The best enterprises today have an active program in place to monitor and manage their web experience.

Defining the Problem

There are two major components to the user's web experience, performance and errors.

Performance is the responsiveness of the application; the time the user has to wait between a mouse click and the arrival of the next page of information.

Errors are when the web server fails to deliver the information, page, video or audio that the user expected. The user is now required to retry the request, find some other way to obtain the information, or give up in frustration.

Errors can be further subdivided into two categories: errors caused by a network fault, and errors caused by a datacenter or browser fault. Network problems can include bad connectivity, missing routes, congestion and/or link failure. These errors prevent the user from communicating with the data center. Datacenter or browser errors occur because of client or server configuration problems, web site errors, software errors, permission errors and many other reasons. These problems are often user or content specific, and are harder to find and fix unless their effect is pervasive.

This report looks at data center and browser errors, along with methods to monitor and diagnose these problems. It also reviews the two primary tools available to assist the enterprise: synthetic test services, and real user monitors.

Managing Site Errors

Seeking out errors and their cause happens in two ways, in a crisis or in a methodical ongoing quality effort. A crisis occurs when some event suddenly causes a major portion of the site to fail, or blocks a group of users from reaching the service or information they desire. The phones start to ring and all hands are immediately applied to finding and solving the problem. Tools applied in a crisis must triage the problem for immediate action.

Quality management, on the other hand, is an ongoing process usually done when there is not crisis at hand. This effort is pursued by one or two individuals who monitor the site looking for problems, and then cure them in a prioritized order. Problems found this way often affect only a small percentage of transactions, and don't merit the attention experienced in a crisis. They can, however, impact users and revenue either by slowing down or discouraging users of the system. The tools need to provide in-depth forensics so that web technicians can sift through details and find failing components.

The Tools

Two types of testing tools are available to aid Enterprise IT Management in their hunt for the cause of errors: synthetic monitoring services, and real user monitors.

Synthetic services use agents placed around the Internet to execute transactions as if they were users. Their synthetic agents run a predefined script specific to the website being tested. This script will execute one or a few typical user interactions, and test the results for performance and the presence of errors. The same script is run at periodic intervals (e.g., once an hour, half hour or quarter hour), and is run on agents placed in the geographies of interest (e.g., major US sites, European sites, or Asian sites)

Real user monitors employ appliances that reside at data centers, and passively monitor the transactions of all users. This information is then sorted and stored in a database, and made available to the web site technician. The data can be sorted and presented in different ways to find classes of users, geographies, carriers, servers, or URLs that are causing errors. The real user monitor shows both real-time and historical information about the actual experience of a web site's users.

A third approach – web logs – needs to be mentioned here. Each web server creates a log of each user transaction. Data mining these logs is possible, but presents a number of problems. Firstly, they do not contain any information about the TCP connection. If a user attempted to connect to the server but was unable to do so, this information is lost. A modern web datacenter has layers of functionality between the servers and the users including SSL accelerators, load balancers, compression devices and caches. Failures in this infrastructure go unnoticed in the web logs.

Secondly, web logs are created by each web server. To get the whole picture of a web site's performance, all the web logs need to be collected, correlated and analyzed. This is logistically difficult and certainly does not provide real-time information.

We assume enterprises depending on their web applications for business will need better functionality than web logs provide. We therefore focus on synthetic services and real user monitors.

Given these two operational modes, crisis and quality management, lets evaluate how well synthetic and real test tools help find and solve web errors.

Crisis Management

A crisis occurs when some change to the network, data center, software, or content causes a significant number of users to be impacted. Detection usually occurs when users start calling the support line. Resources are quickly redirected to diagnose and resolve the problem.

Three steps are required to promptly handle a crisis, detection, diagnosis and repair. Often a crisis is detected through user complaints. It would be preferable to be notified by instrumentation early in the crisis so a problem can be diagnosed and fixed before a large number of users are affected.

The first step in diagnosis is to determine what portion of the infrastructure is at fault, so that the right team can be assigned to fixing it. Is this a network failure, a data center failure, or a problem with content? Drilling down into the failure data should provide this information. Further isolating the problem requires having sufficient diagnostic data to determine which specific users, servers or URLs are causing errors. Resolution of the problem can then quickly follow once the root causes is understood.

Crisis Detection

Detection of a major site problem is a strength of the synthetic service. Testing of the site is continuous, so as soon as the problem occurs, the synthetic service will detect it. How quickly it is detected will depend on the granularity of testing, i.e. on how often the synthetic test service runs its script. Problems that occur during off-hours, such as a major new content update that has errors, will be detected and reported before users start hitting the site in the morning. If properly scripted, the synthetic service will notify web site technicians quickly.

Another advantage of the synthetic service is its ability to test unreleased content. By creating a synthetic script that executes transactions with a test web site, content problems can be discovered before going on-line with a new release.

Real user monitors rely on user traffic. Problems are only detected once users begin to use the site. If thresholds are properly set, notification comes quickly once users arrive at the site. A threshold based on the percentage of errored user hits for instance, detects a problem quickly, even for low user counts.

A significant advantage of real user monitors is that they detect all site problems experienced by users, not just those experiences scripted into the synthetic test tool. If an overnight update has errors in a new section of the web content not tested by the synthetic agent, only the real user monitor will detect those errors.

Crisis Diagnosis

The key to a rapid diagnosis is to have sufficient data to isolate the problem. A single user failure can be the result of a bad browser, connection, load balancer, server, content, cookie, authorization or some combination of these factors. If hundreds of interactions are available to study, the technician can quickly determine what attributes are common among the failing transactions, and thus isolate the issue.

Synthetic test services create a fixed amount of data during a test interval, based on the test script, the frequency it is run and the number of distributed agents involved. Much of this data may not be relevant to the errors being diagnosed if the scripts do not create the right combination of factors.

Real measurement appliances generate data quickly as soon as users start to use the site. If the site failure affects a significant percentage of the users, a large volume of error data quickly becomes available. This error information is directly relevant to the current time, users, browsers, network conditions, server configuration, content release, interest of the users, etc. The accumulated data can be used to quickly isolate the failing components and direct efforts to get it resolved.

Quality Management

During those times when a crisis is not diverting all resources, leading enterprises pursue a quality management process to find and fix low percentage errors on their sites. Quality management is an ongoing process of collecting data, isolating and prioritizing problems, implementing fixes and then repeating the process cycle. Although these problems only affect a small percentage of user interactions, they could be having a significant affect on an important revenue stream or customer segment. They could also bloom into crisis-level problems if the interest or behavior of the larger client base shifts.

Quality is achieved by following a continuous improvement model. First instrumentation is used to find errors. These errors are then sorted into related groups to determine the kind and cause of the errors. The business impact of each group is then determined, and the group or groups having the largest impact are targeted for further diagnosis and correction. Following the corrective changes, the cycle repeats, with another round of error detection.

It may be tempting to ignore small percentage error hits, making the assumption that the vast majority of users are happy using the site. Closer analysis may show that the errors are impacting an important set of users or business partners, or that errors are occurring in a revenue-generating portion of the web site. Without detailed knowledge of the exact nature of these errors, it is hard to know what to fix and if it is worth fixing.

Instrumentation for this process must be able to find all the errors occurring as clients use the site. These errors have to be captured with the details of each transaction so that later analysis can find the correlations between failures, and thus identify the problem. For instance, if the site is responding with errors whenever a particular browser version is being used to read a particular content area, both browser type and URL will be needed to isolate this problem. Errors may be session, temporal, content, user geography, browser or authorization related.

Quality Detection

The key to finding low percentage errors is to watch all the transactions. The synthetic service suffers here because it does not watch actual user transactions, but instead creates its own. If errors are being caused by the web servers or their content, and if the synthetic service is testing that portion of the content, errors will be detected. In other words, errors that are predicted, and then scripted will be found, but unexpected errors will not. Errors caused by a browser setting or by a specific browser type may not be seen as well as those caused by authorization problems or bad cookies.

Web site content is not stable because many enterprise departments are adding or enhancing their portion of the site on a weekly or daily basis. To insure good test coverage, the synthetic scripts should be enhanced in parallel with content updates, but this task often falls low on the priority list. As the web site content drifts away from the script, the value of the testing service decreases.

Short-lived errors can crop up, usually related to network congestion or a failure in the infrastructure. If the synthetic service tests during the short period when errors are occurring, the errors will be detected. If the error appears and disappears between one testing time and the next, it will not be noticed.

One area of strength for the synthetic service is in testing sites with well managed error responses. These sites do not issue an HTTP error, but instead redirect the user to a page indicating the content was not available, and suggesting another path or providing a link to the home page. Even though the HTTP protocol does not see this as an error, because a valid page was displayed, this error is detected by the synthetic service because the script can compare the delivered page to the expected page. A mismatch will be reported as an error.

Total sample size can be an issue for a synthetic service. The low frequency of errors, combined with the low frequency of synthetic test execution may not yield sufficient error samples to get attention. Increasing the testing rate of the synthetic service can help, but the synthetic service soon becomes an additional load on the site.

Lastly, the synthetic service may cause errors of its own. False positives are errors seen and reported by the synthetic service that are specific to its browser, access script or content, and thus don’t indicate a web site problem, just an issue with the test. These false positives tend to reduce the veracity of the synthetic service results, and generate non-productive work for the web support team.

Real user monitors do not have these drawbacks, because they are measuring the actual user experience. Real user monitors are able to detect session errors, momentary errors, content errors, authorization errors, and any other problems that cause the user to receive an error response.

Quality Diagnoses

To determine their impact on the clients or on the enterprise, errors need to be categorized into related groups. Sufficient data and detail in the data is required to determine the different types of errors that are occurring. During the crisis phase we are looking for the one problem that was causing trouble. But during the quality phase we expect there will be a mix of related or unrelated issues, and these issues need to be sorted out so they can be prioritized and their business impact evaluated.

Synthetic services struggle with this step of the process. The data set collected is limited to the browser/cookie/location/content variable chosen in the test script. If these combinations find errors, well and good. But if users are creating other combinations that fail, the synthetic service will not capture them. The ability find and categorize errors is hit or miss.

Errors related to user volume will also go undetected because the synthetic service has no knowledge of user load. Some services now connect to traditional server management tools, and would thus allow correlation to server load. But the synthetic service has no direct knowledge of the number of user requests occurring and how that relates to error rates.

Real user monitors have a complete record of the user transactions including the details of browser, cookie, geography, content and session. With this rich and deep data set it is possible to categorize and correlate errors into workable groups. These groups can then be ranked by business impact, and action plans formed to tackle the worst offenders. Task assignment is clear because the error correlation identifies the network team, data center team or content team based on the common characteristics in each error group.

Table 1 shows a summary of the two modes of operation (Crisis and Quality management) and of the two steps within each mode (Detection and Diagnoses).

Table 1 - Synthetic and Real User Comparison

Phase	**Crisis Detection**	**Crisis Diagnoses**	**Quality Detection**	**Quality Diagnoses**
Synthetic Service	OK	Marginal	Poor	Poor
Real User Monitor	OK	OK	OK	OK

A Real Enterprise Example

Let us look at an example group of web site hits to demonstrate the value of detailed real-time performance information. We analyzed a block of about 1 Million hits on an enterprise data center over about an 11 hour period. The top level statistics are as follows:

Table 2 - Web Hit Statistics

Hits	**Client IP Addresses**	**Server IP Addresses**
1,103,655	32,692	12

Table 3 - Web Hits HTTP Status

HTTP Status	100	200	302	304	404	503	504
Hits	1,542	778,268	123,885	143,604	41,025	2,720	12,345
Percent	0.1%	70.5%	11.2%	13.0%	3.7%	0.2%	1.1%

Other HTTP status codes were received, but were all at percentages much less than 0.1%. We can see from these statistics that the vast majority of hits were successful (100 thru 304 total 94%). Not Found (http status 404) accounts for 3.7%, which may be due to old links or missing content. The 503 (Service Unavailable) and 504 (Gateway Timeout) status codes are server errors, and appear to be only affecting 1.3% of the traffic. This appears to be a small number, but needs further investigation to determine if it is causing users a problem.

Analysis shows that 16.4% of overall requests were for a page named ‘joblist.html’, and this was the most frequently requested page. Of the 180,514 requests for ‘joblist.html’, 1,794 (1.0%) resulted in a 503 error, and 7,249 (4.3%) resulted in a 504 error.

If we then evaluate the client IP addresses we find that 11.3% (203 of 1794) joblist.html requests with a 503 error came from two specific class C addresses, within the same class B address space. This may indicate these addresses are from a common ISP or service provider. We further find that 12% (868 of 7249) joblist.html requests with a 504 error came from these same two specific class C addresses.

Furthermore, the two specific class C addresses had an error rate of 2.6% (503) and 11% (504) when requesting joblist.html, as opposed to error rates of 0.6% (503) and 2.7% (504) when requesting all URLs.

The progressively greater percentage of errors as seen by the users based upon the deeper inspection of the problem is summarized in Figure 1.

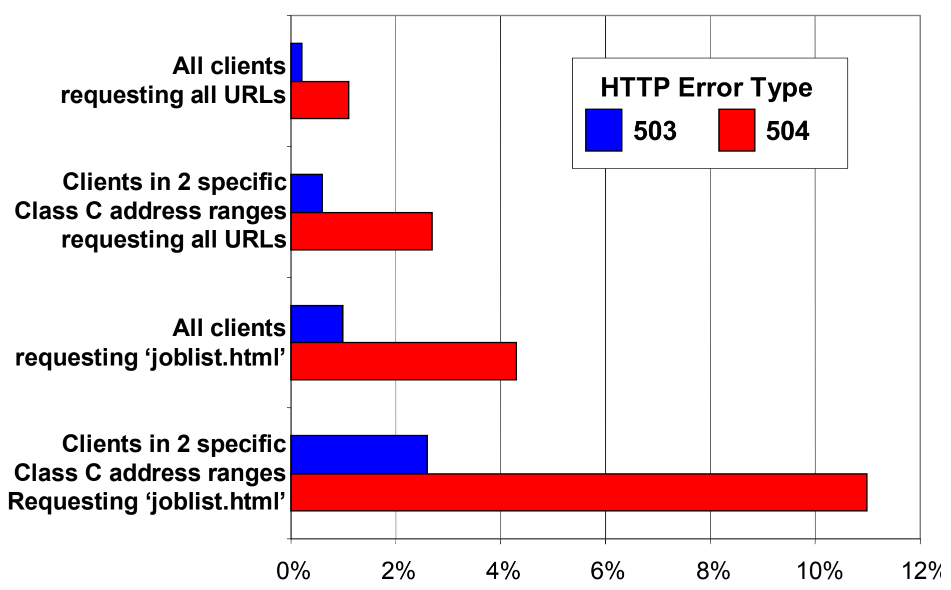

Figure 1 – Finding Progressive Error Significance by Quality Diagnosis

Making the assumption that a small percentage of errors will not substantially affect utilization of the site or service may be inappropriate. If all users are experiencing a 1% error rate, they may well be able to successfully use the site. But here we show that a subset of users are experiencing over a 10% error rate, which makes the site at least frustrating and possibly unusable. If these clients are users valued by the enterprise, remedial measures must be taken to isolate and fix this problem.

Case Review

The case reviewed above clearly falls into the Quality Management area, not crisis management, due to the low percentage of overall hits taking errors. However, the users affected may have significant value to the enterprise. They may be part of a new market being opened, they may be high spenders, or they may be a part of the enterprise management team. Finding these affected users and clearing up the cause of their frustration may be important.

We still don't know the root cause of this specific problem. If the problem is specific to the joblist.html web page, a synthetic tester may uncover the problem, but would not raise it to the level of concern shown in Case 4 above. The synthetic tester would only show the problem if it is configured to test that specific page in the specific way that users are using when the problem occurs.

If the problem is related to a particular network path into the data center, as may be indicated by the specific addresses being affected, the synthetic tool would not see the problem at all, unless it happens to be in one of those specific class C address spaces. If the problem is related to the specific browser type and version that the clients are using, again the synthetic service will not spot the problem.

Lastly, the problem may be occurring due to a combination of all of the above factors. Perhaps packet loss on the ISP supporting those specific clients is exacerbating a data center problem on the specific server that supports 'joblist.html'. The real user test appliance is the only solution that will find and isolate these combinations quickly so that corrective action can be taken in a timely manner.

Recommendation

Any organization serious about the quality of the web experience they provide to their users should strongly consider the use of a real-user measurement appliance. While synthetic services can provide analysis for a crisis situation, the job of delving into lower level problems, isolating and then correcting them to improve service delivery to all clients can be done most effectively with a real-user measurement tool.

John Bartlett is Vice President of NetForecast, and has 24 years of experience designing at the chip, board, system and network levels, with a focus on performance. John led the team that built the first VLAN implementation, one of the first ATM switches, and he is a leading authority on the behavior of real-time traffic on the Internet. He can be reached at john@netforecast.com.

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.

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

Additional information is available at: www.netforecast.com