

Averages Hide the Real End-User Experience: Apdex Tells the Full Story

By Peter Sevcik
April 2007

In February 2007 NetForecast conducted an extensive survey of APM (application performance management) best practices¹. One of the many findings was a strong correlation between good communication about application performance and actually delivering good performance. The top group excelled at communication relative to the other best practices. Accordingly, this column explores various ways to communicate information about application performance.

First we must start with application performance measurements. NetForecast has been conducting ongoing measurements of the Apdex Alliance Contributing Members' websites (there are currently 10 members). We are using a synthetic agent service that gets the home page of each website from various locations around the world every 15 minutes. For the purposes of this article, we are limiting the discussion to the data gathered from 5 locations in the U.S.: California, Colorado, Florida, Minnesota and New York.

The data was grouped so as to track the overall performance that users in each of the states would experience as they tried to access the 10 sites. The tests simulate a business-to-business e-commerce type of interaction. The synthetic agents simulate such a user because they are generally well wired to the Internet, with access lines of 100 Mbps.

Readers should not focus on the actual 10 websites, the testing service used for the experiment, or even the choice of the 5 "user" locations. The quality of the experiment is sufficient to have a good data set that represents Internet performance over the three week period of February 25 to March 17. The purpose of this article is to discuss what one can now do with the data set.

We developed three types of reports which represent three distinct ways by which the person developing the report interacts with his enterprise.

Deviation From Normal

The simplest way to report the data is to not have any conversation with the application users or management. Let the data report on itself. The premise is simple. Figure out what is the typical response time for each region, and then report days that were significantly slower than normal as poor performance. Some measurement tools have this feature built in as a standard report.

We start by averaging all the measurements per region per day. Then we perform a five-day rolling average of daily performance in each region. The algorithm next adds 50 percent to the rolling average, which becomes the good/poor performance threshold. Each day's average is then compared to the automated threshold.

The result shown in Figure 1 is the delta from the threshold for each day, where negative values are the time in seconds that the day has performed below the threshold.

For example, if the recent rolling average is 4 seconds, then the threshold is 6 seconds (i.e., increasing the 4 second average by 50 percent, as explained above). If today's average performance was 5 seconds, then the result shows that the day was 1 second

NetForecast Report
5086

©2007
NetForecast, Inc.

¹ "Application Performance Management: Best Practices Do Work," by Peter Sevcik and Rebecca Wetzel, *Business Communications Review*, May 2007

below the threshold, or negative-1 second. Below-zero scores are good; scores that are above zero are indications of poor performance relative to the norm.

Figure 1 shows that over the three weeks, performance was generally good across all the regions except for three specific incidents. Florida saw poor performance on March 2 and March 12 while California was hit on March 8.

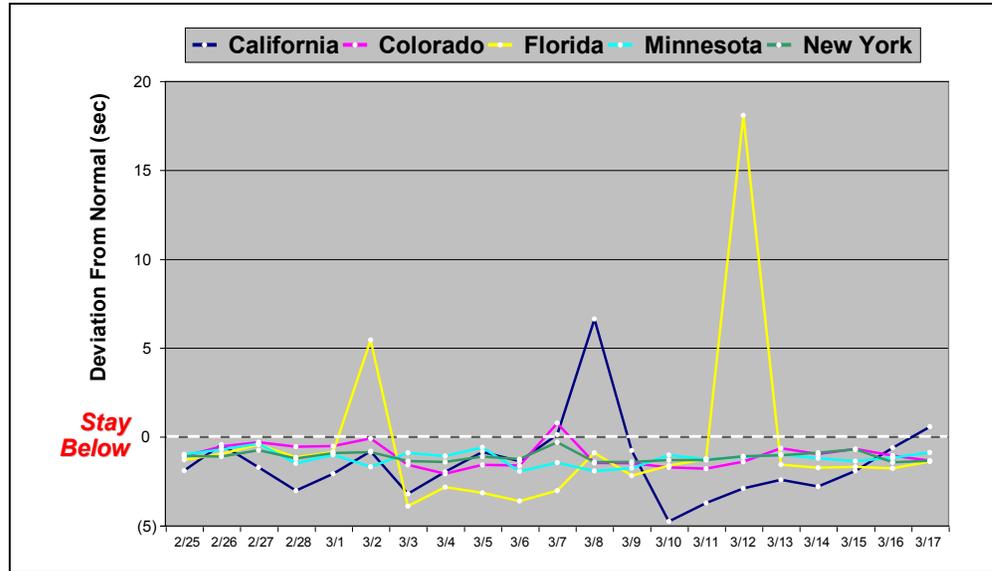


Figure 1 – Average Deviation From Normal

Average With Threshold

Let us now become more proactive, engaging some members of corporate management in the process, and figure out a specific threshold that represents the good/poor performance boundary in this scenario. In this case, we applied the target of 4 seconds. It turns out that on a good day, the average from all regions to all sites is just over 2 seconds. The average for all days (good and bad) is 2.9 seconds.

So the results are what one would generally see in a B-B type of interaction; several organizations have recently stated that such interactions should stay within 4 seconds. Let us assume that the person who is generating these reports has an informed conversation with the business manager who cares about this application, and they agree on a 4 second response time target.

Figure 2 shows the exact same averages per region per day as used in the first report. The agreed-upon target of 4 seconds is clearly marked. Any region-day below the 4 seconds is operating within the service level agreement.

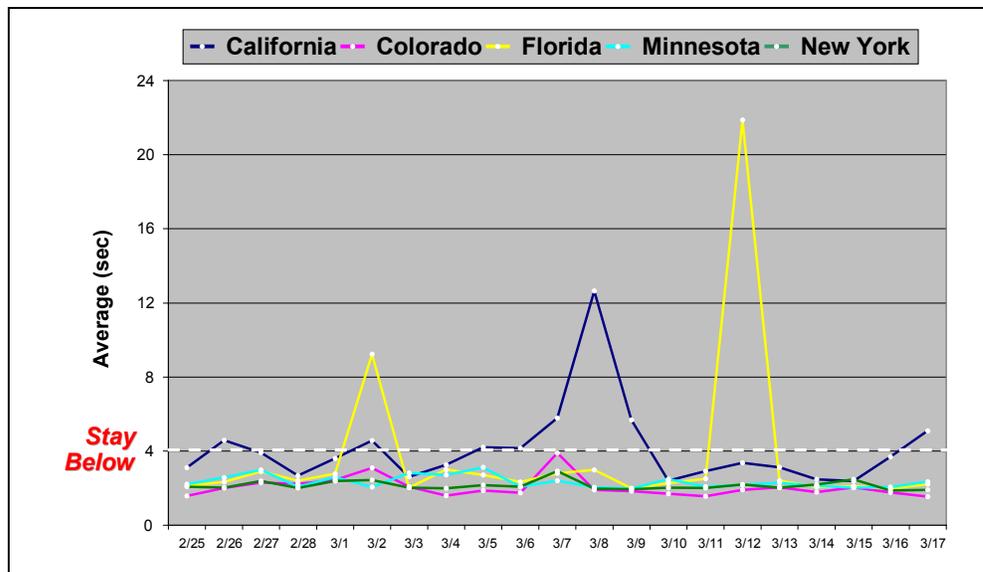


Figure 2 – Average With Threshold

In this case, the result is remarkably similar to Figure 1. The same three days show poor performance being delivered to the same regions. The only new information is that California goes a little above the threshold a few times.

Apdex With Objective

The more sophisticated dialogue between the IT staff and business managers would be to apply the Apdex methodology (see www.apdex.org). Here the conversation has two parts.

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.

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:

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

Step 4 – Show Results

Show the Apdex Index value (on a scale of 1 to 0) together with the corresponding threshold T.

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

First there needs to be an agreement on the Apdex performance target called T. T is the threshold between satisfied and tolerating performance as seen by the user. Let us also assume that the same 4 seconds agreed to above is used here as well.

Now the second part of the conversation. Given that the performance when reported by Apdex will have a range of 1 (excellent) to 0 (disaster), what should be the performance objective? The Apdex specification says that an Apdex score of 0.94 to 1.00 is excellent, while 0.85 to 0.94 represent good performance delivery. Let us assume that the new service level agreement using Apdex is for a target of 4 seconds delivered with an Apdex score of 0.90 or better.

Figure 3 shows the exact same data gathered over the three weeks using the Apdex methodology. In this chart, better scores are up rather than down as in the previous charts. So the poor-performing region-days are those below 0.9

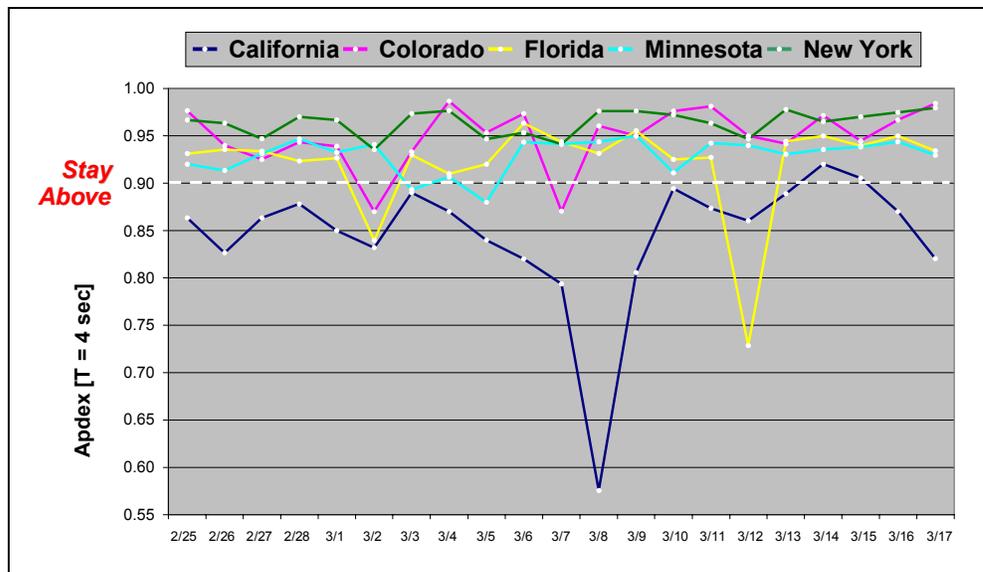


Figure 3 – Apdex With Objective

This view of the same data shows strikingly different information. The first observation is that the users in California are seeing chronically poor performance.

This is a perfect example of getting phone calls from the users or managers in California complaining about performance. If the IT department were using either of the previous two averaging methods of reporting, they would have shrugged off the complaints with, “I don’t know what they are talking about, it all looks normal from here.” But if they were using Apdex reporting, then they would have been working on the California problem already, since it is clearly below the SLA.

The other interesting result is that while the two averaging methods both reported the worst incident as Florida on March 12, the actual worst incident, as reported by Apdex, was California on March 8. The performance in California on March 8 was so bad that the Apdex score dropped to nearly 0.5 which is considered unacceptable! This is a completely different conclusion than found by the averaging methods. Furthermore, Colorado and Minnesota saw some poor performance days that went undiscovered in the averaging methods.

Table 1 shows the three approaches’ ability to find poor performance days. Deviation from Norms found 6, Averages with a Threshold found 10, and Apdex found 25. Apdex found more than twice as many performance incidents than the averaging methods. Apdex finds more of them because it finds the “tolerating” and “frustrated” user sets. Averaging washed these groups away.

Table 1	
Region-Days With Performance Incident	
Deviation From Normal	6
Average With Threshold	10
Apdex With Objective	25

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

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

Additional information is available at: www.netforecast.com

NetForecast and the curve-on-grid logo are registered trademarks of NetForecast, Inc.

How The Reporting Method Affects SLAs

Let's look at the difference when the deviation from normal and the Apdex approaches are used to drive a service level agreement (SLA). Figure 4 shows the percentage of region-days during which application performance met the following two SLAs:

- The average within each region must be less than a 50% deviation from the normal performance for that region.
- The Apdex score based on a T of 4 seconds must not fall below 0.90 for each region.

As expected, the averaging method shows a more optimistic view of performance compare to the Apdex view over a 10 week reporting period. However, the real user experience is captured by the Apdex curve and not the deviation from normal curve. The difference between the two shows the dramatic under-reporting of the true performance picture. This reporting gap reflects fundamentally different viewpoints between IT and business managers.

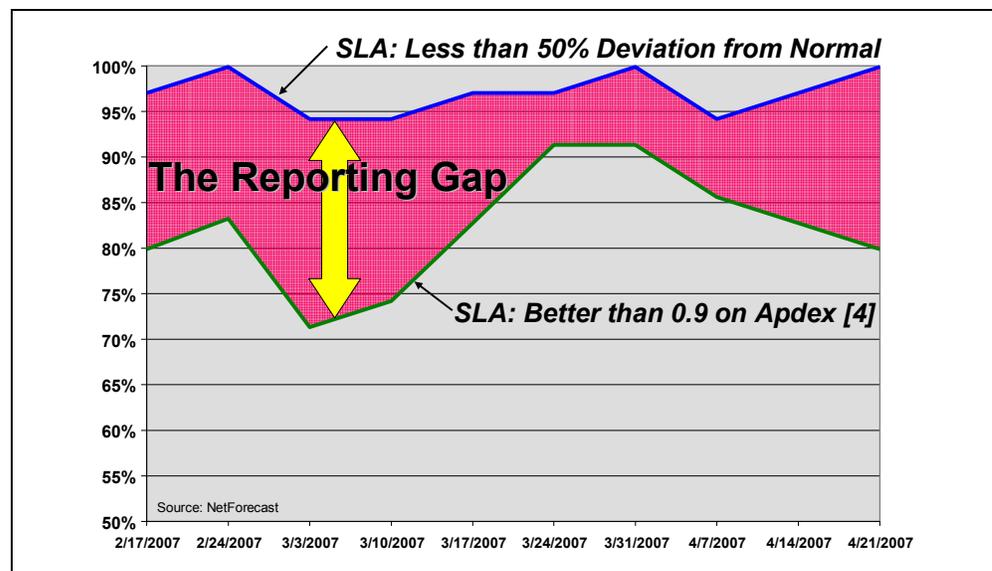


Figure 4 - Region-Days Meeting Their SLA (%)

Conclusion

The results clearly show that if you want to learn what is really going on with user performance, you should process numbers through the Apdex formula. And if you want a meaningful dialogue with business managers about application performance that is linked to business needs, you should use the Apdex methodology as a common frame of reference.

The Apdex approach is free for the using. Anyone can export their numbers into an Excel spreadsheet and produce the charts shown here for their network.

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.