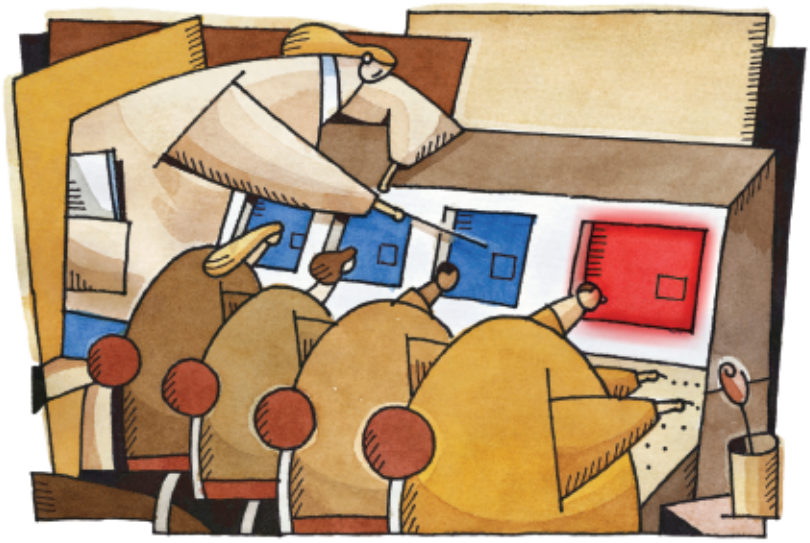
Network and Application Performance Alarms – *What's Really Going On?*



Jim Metzler
Ashton, Metzler & Associates
jim@ashtonmetzler.com

Introduction

The June 2005 IT Impact Brief was entitled "It's the Application Stupid". That brief established the business relevance of infrastructure and applications. In particular, that brief contained market research that demonstrated that a company's business managers are far more likely to see the value provided by the IT function as coming from applications than coming from the IT infrastructure.

I believe that on an ever-increasing basis the most common technique that infrastructure and management organizations will use to demonstrate their business value is to show what they do to enable the company's key applications and business processes. Based on that belief, several recent IT Impact Briefs have discussed what it takes to support applications. This includes the last brief that discussed the need to have multiple sources of management data.

This goal of this impact brief is to present a reality check on how companies currently deploy and respond to network and application performance alarms. To provide input to this brief, a survey was given to the NetScout community. Given that two hundred and seventy IT professionals responded to the survey, the survey responses give us a very accurate look at the use of network and application performance alarms. Throughout this brief, the respondents to that survey will be referred to as The Survey Respondents.

In order to gather additional insight, I interviewed two of The Survey Respondents. One of the interviewees is an IT project leader for a company in the transportation industry. The other interviewee is an IT team leader for a company in the medical industry. Throughout this brief, the interviewees will be referred to as The Project Leader and The Team Leader.

Are IT Organizations Setting Alarms?

We asked The Survey Respondents if their organization sets static threshold performance-based alarms. Roughly three quarters (72.8%) of The Survey Respondents indicated that they do. The Survey Respondents were then asked to indicate the network and application parameters against which they set the alarms. Their answers to that question are contained in Table 1. Note that The Survey Respondents were instructed to indicate as many parameters as applied to their situation.

Table 1:
Percentage of Companies that Set Specific Thresholds

Parameter	Percentage
WAN Traffic Utilization	81.5%
Network Response Time (Ping, TCP Connect)	58.5%
LAN Traffic Utilization	47.8%
Application Response Time (Synthetic Transaction Based)	30.2%
Application Utilization	12.2%
Other	5.9%

As is shown in Table 1, the vast majority of IT organizations set thresholds against WAN traffic utilization or some other network parameter. Less than on third of IT organizations set parameters against application response time.

The Project Leader indicated that his organization currently sets performance alarms only on WAN traffic utilization. However, he is currently researching how to offer their internal users a Service Level Agreement (SLA) at the application level. Offering an SLA at the application level will require them to set performance alarms on application response time.

The Team Leader stated that his company benchmarks all major new applications before they are deployed in order to understand how network parameters such as delay and packet loss impact the performance of those applications. His organization also sets performance alarms on WAN utilization, network response time, and application response time. They use the information gathered from the alarms combined with their knowledge of the applications to ensure that the applications are performing as expected.

Setting thresholds against network utilization is clearly a good thing to do. Many companies use the rule of thumb that says that you should not let network utilization exceed seventy or eighty percent. The reason for this rule of thumb is that once the network utilization reaches those levels the network begins to experience congestion that usually leads to packet loss. Packet loss causes some applications to perform poorly.

However, companies that manage network and application performance based just on network utilization are implicitly making the following two assumptions:

1. If the network is heavily utilized, the applications are performing poorly
2. If the network is lightly utilized, the applications are performing well

The first assumption is often, but not always true. For example, if the company is primarily supporting email or bulk file transfer applications, heavy network utilization is unlikely to cause unacceptable application performance.

The second assumption is often false, as it is quite possible to have the network operating at relatively low utilization levels and still have the application perform poorly. An example of this is any application that uses a chatty protocol over the WAN. Chatty protocols, such as CIFS (Common Internet File System) or NFS (Network File System), exchange tens or even hundreds of messages between sender and receiver for each transaction. These protocols were designed to run over a LAN and typically perform well in that environment. However, when an application that uses a chatty protocol is run over the WAN, it is likely to perform badly even if there is extremely light WAN utilization. A good example of this is when an IT organization consolidates Microsoft servers in a centralized data centers. Since many Microsoft applications use CIFS, this means that CIFS is now running on the WAN between the company's branch offices and the centralized site.

The Survey Respondents were also asked to indicate the approach that their company takes to setting performance thresholds. Their answers are contained in Table 2.

Table 2: Approach to Setting Thresholds

Approach	Companies
We set the thresholds at a high-water mark so that we only see severe problems	64.3%
We set the thresholds low because we want to know every single abnormality that occurs	18.3%
Other (Please specify)	17.4%

Of The Survey Respondents that indicated other, their most common response was that their company sets the thresholds at what they consider to be an average value.

One conclusion that can be drawn from the data in Table 2 is that the vast majority of IT organizations set the thresholds high to minimize the number of alarms that they receive. While this approach certainly makes sense from an operational perspective, it does mean that in most IT organizations the majority of the alarms are ignored.

The Project Leader provided a somewhat different reason for setting thresholds at a high-water mark. He stated that his organization primarily uses the performance alarms as input into their capacity planning process. Based on this usage of performance alarms, it makes sense to set thresholds at a high-water mark.

What happens to the Alarms?

The Survey Respondents were asked to indicate the approach that their organization takes to responding to the alarms that they do receive. Their responses are contained in Table 3.

Table 3: Responding to Alarms

Approach	Percentage
We try to look at the alarms regularly and do something with most of them.	32.2%
We look at every alarm and determine what to do with it.	28.0%
We utilize the rules programmed into our tools to determine the severity. If it hits the priority category, someone is beeped.	26.8%
No one does anything with them	9.6%
Other	3.3%

The data in Table 3 highlights the fact that responding to alarms is largely a manual task. In particular, only about a quarter of The Survey Respondents indicated that they use rules programmed into their tools to determine the severity of the alarms. The fact that roughly one IT organization out of ten indicated that they do nothing with the alarms is further indication that the current approach to setting and responding to performance alarms needs to be significantly enhanced.

The need to enhance the current approach to setting and responding to performance alarms was reinforced by the answers (Table 4) The Survey Respondents gave to the question of "Who in your organization normally first recognizes performance problems?"

Table 4: Performance Problem Recognition

Problem First Recognized By:	Percentage
End user calls help desk	40.7%
NOC Operator	31.6%
Network Engineer	19.6%
Application Support	4.7%
Other	3.3%

The issue that is highlighted by the data in Table 4 is that despite the attempts that IT organizations have made to become more proactive, end user calls to the help desk are still the most common way of identifying a performance problem.

The Project Leader stated that if the problem is an outright outage, that the NOC is as likely as the end user to be the first to recognize the problem. However, he also stated that if the problem is application degradation, then in almost all cases it is the end user who first recognizes the problem.

The Team Leader stated that in their larger facilities, a transient event such as a user downloading a large file is unlikely to impact other users. As a result, in their larger facilities the NOC will most likely be the first to notice a problem. However, The Team Leader also stated that in their smaller facilities, it is common to have a transient event cause an end user to report a trouble.

Conclusion

IT organizations as a whole are under continued pressure to demonstrate the business value that they provide to the company. One of the best ways that infrastructure and management organizations can show business value is by ensuring the performance of the company's network and applications.

Setting performance thresholds clearly helps IT organizations manage the performance of the company's network and applications. However, the research data presented in this brief highlights the fact that there is a management gap between the current approach to using network and application performance alarms and what is required to proactively manage a network and applications. Some of the indicators of that gap are:

- Less than a third of companies set a performance threshold on applications
- Two thirds of companies ignore the vast majority of alarms
- Only a quarter of companies utilize rules programmed into their tools to analyze the severity of alarms
- The most likely person to recognize that there is a performance problem is an end user

The next IT Impact Brief will discuss Voice over IP (VoIP). This discussion will include the breadth and extent of VoIP deployment, as well as the efforts that IT organizations are taking to manage this highly-visible, delay-sensitive application.