Beyond Monitoring – Root-Cause Analysis

WHITE PAPER

With the introduction of NetFlow and similar flow-based technologies, solutions based on flow-based data have become the most popular methods of network monitoring. While effective, flow-based network monitoring has a significant limitation – it lacks the depth of data to perform true root-cause analysis. This paper introduces a new concept – Network Monitoring for Analysis – a technique that lets you know when there's a problem and why there's a problem, and illustrates a solution that provides the uncompromising performance and capabilities needed to address Network Monitoring for Analysis.

WildPackets, Inc.
1340 Treat Blvd, Suite 500
Walnut Creek, CA 94597
925.937.3200
www.wildpackets.com

Beyond Monitoring – Root-Cause Analysis

Beyond Monitoring – Root-Cause Analysis

Introduction

The network monitoring landscape was forever altered by the introduction of flow-based data. This trend began with the introduction of NetFlow by Cisco, and quickly accelerated as other network equipment vendors either adopted this proprietary Cisco standard or implemented similar technologies. NetFlow and similar technologies enable the generation and export of network performance data based on network flows, a unidirectional sequence of network packets that share certain characteristics, including the source and destination IP address, the source and destination port, the IP protocol, the ingress interface, and the IP Type of Service. Standard network equipment can be configured to generate flow-based data, compile the resulting flow records, and export this information to third-party collectors for processing and analysis. Flow-based technologies quickly replaced SNMP as the core technology used for network monitoring as flow records contain more detailed network information than SNMP and data collection is far more efficient, and the information still "comes for free" from any supported switch or router.

However, as with SNMP, flow-based network monitoring does not do a complete job. Real-time, flow-based statistics provide visibility into how the network is operating, and can even allow some extrapolation as to how the network will continue to perform, but when conditions begin to degrade the best they can do is to raise the red flag. They provide very little information to determine the root cause of the issue, making it extremely difficult to isolate the problem and implement a permanent fix. This is the key difference between a Network Monitoring and a Network Monitoring for Analysis solution. In Network Monitoring for Analysis the same flow-based statistics are available for real-time network performance monitoring as in flow-based solutions, but detailed network data are also archived for forensic, or post-incident, analysis, eliminating the need to recreate the often ephemeral, anomalous condition that created the bottleneck and providing all of the detail needed to perform root-cause analysis and to address the issue once and for all.

Business Benefits of Root-Cause Analysis

The business benefit of root-cause analysis is clear. Without it you cannot fully identify and address the conditions causing a network problem, whether performance, security, compliance, or process related. The network has become the lifeline for your business, so network problems are business problems, and every minute that the network isn't operating as designed generates increased costs, loss of revenue, or both.

The real question centers on how best to achieve root-cause analysis. With flow-based network monitoring, an additional investment must be made in technology that enables root-cause analysis, as well as the additional management, training, and maintenance expenses that accompany multiple technologies. The real benefit is in deploying a single solution that provides complete network monitoring and root-cause analysis, all in one package, and all from a single vendor.

Difference between Monitoring and Monitoring for Analysis

Many network monitoring solutions are available. All will give you the health of the network, and most will alert you when a problem occurs. However, most network monitoring solutions do not enable you to conduct root-cause analysis.

With Network Monitoring, you have some useful real-time data. With Network Monitoring for Analysis, you have enough data available so you can drill down and analyze when you need to. With Network Monitoring, you know there's a problem. With Network Monitoring for Analysis, you know there's a problem and ***why*** there's a problem.

To use Network Monitoring for Analysis, you need to choose a network monitoring approach that collects the appropriate data.

Approaches to Network Monitoring

Networking monitoring involves three primary approaches:

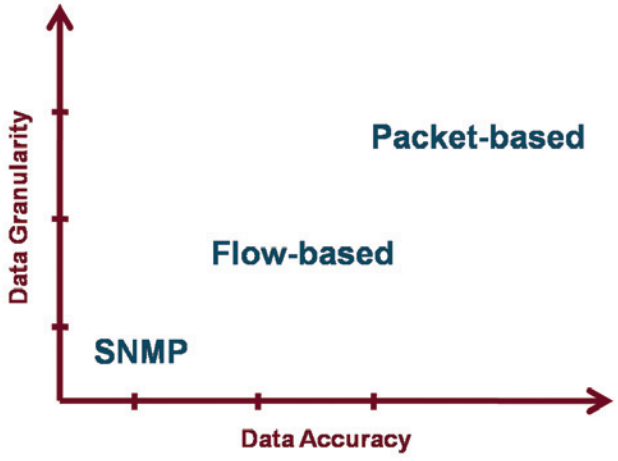
- **Simple Network Management Protocol (SNMP)** is useful for identifying and describing system configurations. It monitors network-attached devices for basic high-level conditions such as up/down, total traffic (bytes, packets), and number of users. Unfortunately it uses polling which has a heavy bandwidth impact as lots of polled information traverses the very network you're attempting to monitor.
- **Flow Records** are the default elements used in centralized, flow-based network monitoring. A "flow" is a sequence of packets that has 7 identical characteristics – source IP address, destination IP address, source port, destination port, layer 3 protocol type, type of service (TOS) byte, and input logical interface. Flow records vary by overall standard, vendor, and configuration. The most common are NetFlow, IPFIX, sFlow, and JFlow. Unlike SNMP, flow-based data yield more detailed statistics and provide good information about the overall health of the network. Flow-based analysis can impact network performance as it relies on the same equipment used to control network traffic – the routers and switches themselves – and can cause conflicts for processing power and memory. Typically, when network conditions begin to tax a switch or router it will revert to its "prime directive" – routing packets - reducing the reliability of the network monitoring data.
- **Packet-Based** captures each packet, using software and/or computer hardware, as traffic passes over a digital network or part of a network. Captured packets are then decoded and analyzed according to the appropriate Internet Engineering Task Force (IETF) RFC (Request for Comments) or other specifications. Unlike flow records which rely on statistical sampling, packet-based approaches generate 100% accurate information for each flow. Also, unlike both SNMP and flow records, there's minimal network impact as all analysis is done locally at the point of capture, and on hardware that is not part of the network routing infrastructure.

TABLE 1: Comparison of Flows

NetFlow	IPFIX	sFlow	JFlow	OmniFlow
• Developed by Cisco • Proprietary • Transit traffic & terminated traffic • Detailed info for each flow • NO payloads • Sampled option not 100% accurate	• Internet Protocol Flow Information eXchange • Emerging IETF standard • Based on NetFlow • Detailed info for each flow • NO payloads	• RFC 3176 • Statistical time-based sampling • Higher speed networks • Much less common than NetFlow • NO payloads • Sampled – not 100% accurate	• Developed by Juniper • Proprietary • Similar to NetFlow • Detailed info for each flow • NO payloads • Sampled per global rate – not 100% accurate	• Developed by WildPackets • Proprietary • Analysis of every packet AND payload • Unrivaled info for each flow • Layer 3 - 7 • 100% accurate • Monitor AND troubleshoot

Regardless of which approach you chose, you're going to have to make compromises. Two metrics that are useful when making those compromises are Data Granularity and Data Accuracy. *The compromises you make here determine whether or not you're able to monitor for analysis.* Data granularity is essential for performing root-cause analysis. The data must include the detail required to reconstruct what's happening at the time the network anomaly occurs. Only packet-based solutions provide the appropriate level of data granularity. Data accuracy refers to how much of the data is included in the generation of key performance indicators on the network and how accurate these calculations are. Both SNMP and flow-based solutions rely on sampling for gathering the data used in network performance calculations. Sampling, though often sufficient, can generate misleading results, especially if a significantly large flow happens to be missed by the sampling algorithm. Again, packet-based solutions analyze 100% of the data on the network, providing the most accurate solution available.

FIGURE 1: Comparing Data Granularity and Data Accuracy across Network Monitoring Approaches



Overhead and cost also need to be considered when trying to determine the optimal solution. Overhead comes in several forms, including increased network traffic due to the monitoring solution itself as well as resource loading on key network devices. Both SNMP and flow-based solutions leverage the processing and storage resources of the same equipment being used to route the packets. When conflicts occur, the integrity of the network monitoring data is at risk. Both also send significant bursts of traffic at regular intervals from the network infrastructure devices to the data collection and processing systems, placing an additional load on what may already be a relatively taxed network. Packet-based solutions use dedicated hardware for both data collection and analysis, eliminating both resource and network overhead.

Cost is also an important factor. Since SNMP and flow-based solutions rely on the existing network infrastructure to generate data, the cost for these solutions is easy to control. Packet-based solutions require additional hardware which often implies a higher cost, but this is typically offset by the gains in deploying a single solution for BOTH network monitoring and root-cause analysis, or Network Monitoring for Analysis.

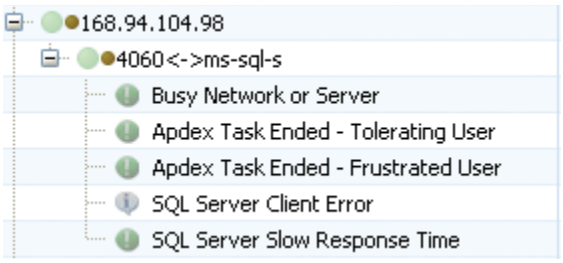
Importance of Packets and Payloads

Often the help desk receives a call where a *particular* user is having a problem with a *particular* application. This might go unnoticed with a flow-based approach as the high-level alerts you've configured may not bring this to your attention. In this circumstance, having the packets and the payload can make all the difference.

Consider this example. Figure 2 shows an excerpt from a packet-based flow analysis from the client side of the flow. Here you can see the port the user was communicating over and that this user is accessing a SQL-based application (4060 <-> ms-sql-s). The following five lines – the result of Expert Analysis – sheds insight into what the user is

experiencing. Just by looking at the flow analysis you know that the user is waiting seconds, if not tens of seconds, for responses between the application they're using and the server.

FIGURE 2: Flow-based client analysis

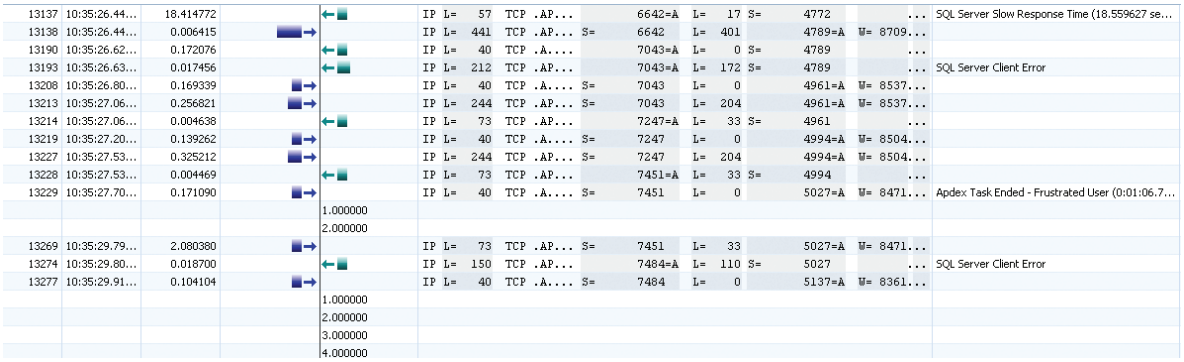


When all packets are captured, Expert Analysis can be done which enables you to monitor the end user's experience.

Figure 3 shows a visualization of this communication – the flow between the server and the client packet by packet. The color bars (blue and green) indicate the source of the traffic and the relative size of each packet. In this example, you see that a request was made and it took 2.080380 seconds for a response to come back. A little later in the communication, you see that a request has been made and no response has come back, even after 5 seconds. The column to the right explains why – SQL Server Client Error.

Without the packets, you wouldn't know the wait times and the response times the user was getting. This is why capturing packets is important.

FIGURE 3: Packets tell you the wait times and the response times experienced by a user.



With a ladder diagram you can see that response times are poor. Coupled with the payload information, you know there was a deadlock condition at the database that prevents the user from completing the transaction.

OmniFlow, in contrast to NetFlow, IPFIX, sFlow, and JFlow, enables the analysis of every packet and payload. So what does this mean? The packet payload is typically the linkage between networking information and application information.

Return to the previous example. One layer down at the payloads you can see the entire payload conversation for the flow. Figure 4 shows the payload for the packets that were highlighted in blue in Figure 3. The user in this example is unable to get into the database and complete their task.

FIGURE 4: Payloads, in addition to packet headers, give you the complete picture.

```
................
..Your server command (process id 169) was deadlocked with another process and has been chosen as deadlock
victim. Re-run your command
.CS01ES............
```

With the payload information, you know there was a deadlock condition at the database that prevents the user from completing the transaction.

Without this information you might resort to trial and error with the user to find out what's wrong, most likely assuming it's a network issue, and never involving the application engineer in the analysis.

With the payload information, you're able to determine not just that there's a problem, but exactly what the problem is – an application design issue that requires the attention of the designer. When you have the payload information, you don't have to spend time reproducing the problem. You already have the packets that illustrate the problem and you're able to dig in and do the analysis. The payload information completes the story and tells you what's at fault – the network or the application.

Value in Archiving Packets

Organizations, whether formally or informally, have been performing some level of post-incident or forensic analysis for years in response to security attacks. An Intrusion Detection System (IDS) or Intrusion Prevention System (IPS) monitoring the corporate firewall will partially process data in the event of an attack, and notify you of the situation, but the data is incomplete since the IDS/IPS is not designed to archive the packets. Determining the attack fingerprint will be difficult, if not impossible, as you have no way of recreating the events. This lack of data for analysis limits your ability to prevent future attacks.

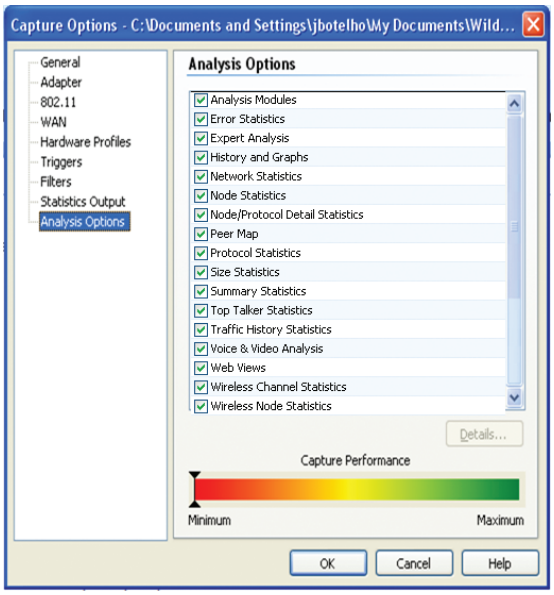
A more efficient method of forensic analysis is to employ data recorders in line with the IDS/IPS to archive the packets. In this case, an attack is recorded by the network data recorder, and the IDS/IPS is triggered. Now you have a complete recording of the entire attack. You're able to analyze the event to reveal the attacker, the method employed, and any damage that occurred.

Typically it is very difficult to prevent a zero-day attack, regardless of the claims made by many security solutions. If the attack has never been seen before no one is going to know its fingerprint. If you're taking a classic security approach you might be alerted to the intrusion, which is good. If you're using network data recorders in line with your IDS/IPS, you're able to go back to the packets you archived and recreate the attack. With this recreation, you might see the IP addresses the attack came from or unique strings within some of the packet combinations that allow you to find repeat attempts of that particular attack, as well as assess the damage done by the original attack, a critical element in meeting today's new compliance standards.

Default Analysis Options

Regardless of whether you run real-time analysis or a combination of real-time and post-incident analysis, it is imperative to determine what data is essential to your analysis needs. What is it that you're looking for? This is important because most software will have all analysis options turned on by default.

FIGURE 5: Determine what you are interested in and disable the other options to increase performance.



While the defaults ensure that you're not missing anything, you're impacting performance by performing superfluous analysis. Defaults are optimized for analysis – not throughput. For real-time analysis, take a look at what the software is analyzing and disable what you don't need. For example, you might not be interested in VoIP or Video. In that case, you can turn off the Voice & Video Analysis option.

NOTE: If you're only doing post-incident analysis, turn off all analysis options. This will significantly increase overall performance. With the analysis options off, the network data recorder will write the packets to disk during the first pass and not do the detailed analysis. When you connect to the recorder and need to review the packets to reconstruct a network anomaly, turn the analysis options you care about back on. The network analyzer will then review the packets and you'll see the results of the analysis.

Placement of Monitoring Points to Enable Root-Cause Analysis

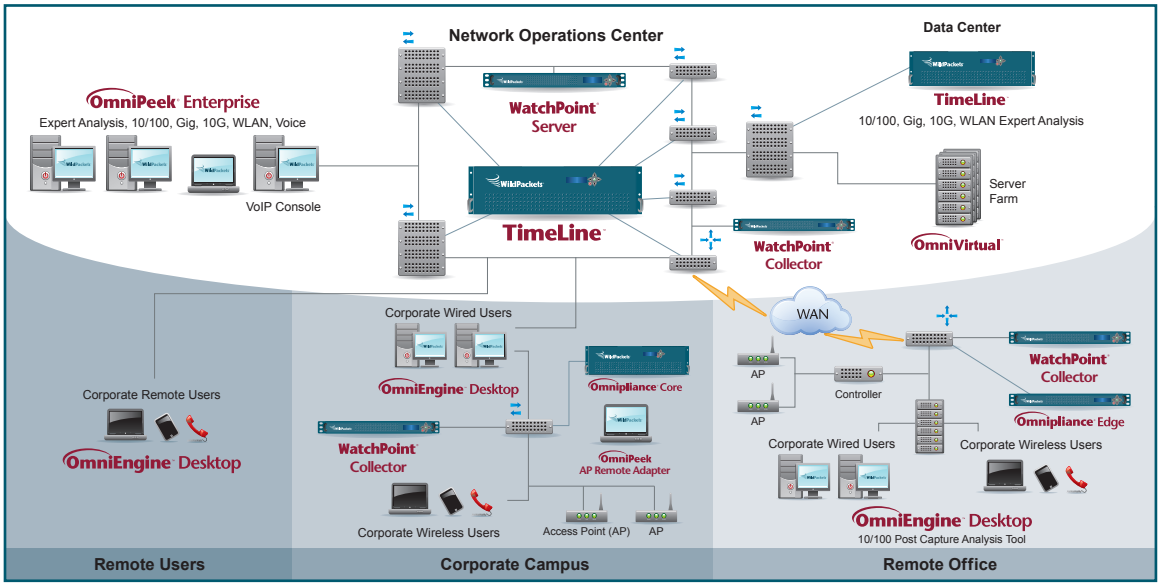
Before you start assessing how and where you're going to deploy the monitoring points in your topology, ask yourself the following:

- Do you want to be able to perform network forensics or post-incident analysis?
- Are you interested in monitoring application performance?
- Are you trying to monitor Voice over IP (VoIP) and Video?
- Is NetFlow and sFlow important to you?
- Do you have virtual servers or virtual applications in your infrastructure?
- Are there 10 Gigabit (10G) segments?
- Do you have wireless in your network?

How you answer the questions affects the type of hardware you deploy as well as the measurement points.

Figure 6 shows a sample distributed environment. In this example, we have a centrally located Network Operations Center (NOC) that's not at the corporate campus, as well as a Data Center and Server Farm. In this example, this is one of the points we want to be monitoring. As there's typically little wireless at the NOC, we probably don't have to worry about monitoring wireless there. We will however have Gigabit and 10G segments to consider.

FIGURE 6: A Typical Distributed Environment



In this topology, the NOC is critical – all packets going out over the Internet funnel through it. We may want to think about placing a larger appliance here, such as a TimeLine network recorder, so that not only can we capture and analyze all of the packets, but store those packets as well. By storing the packets, if there's a problem in the network – whether it occurred a few minutes ago or several hours ago – we still have the packet files available to replay that incident and conduct analysis right from the start. For this reason, we may also want to place another larger appliance in the Data Center to ensure we capture all of the local data center traffic.

Now let's take a look at the Server Farm. Here we might have virtual applications and/or virtual servers. To troubleshoot these virtual applications and virtual servers, we'll probably want software that allows us to capture and analyze virtual traffic – packets that never traverse a real network interface card (NIC), switch, or router. For example, all traffic between an application server and the database server may be resident on the same virtual machine. The packets never traverse a physical NIC, so we would never see this data. We may not even see this with NetFlow or sFlow as the packets aren't going out through a switch or router. Software, such as WildPackets OmniVirtual VMWare Probes, enables us to see the data from within the virtual machine.

Outside the NOC, Data Center, and Server Farm, in this topology, we have remote users, a Remote Office, as well as the Corporate Campus. Each probably uses different technologies for networking. With remote users, many are wireless. We may want to capture data from their desktop or over the wireless network. Capturing wireless data is a very local type of event; it's dependent on point of presence – we need to be within a 300-foot radius of the access point (AP). To capture this type of data we'll need remote sensors or the ability to turn an AP into a sensor.

At the Remote Office, there might be a lot going on. We know for sure everything from that Remote Office funnels through the NOC and Data Center. There's no need to have an appliance for storing packets locally. We may want to have an appliance, such as the Omnipliance Edge, that allows us to monitor and analyze up to Gigabit traffic.

Finally, at the Corporate Campus, we'll probably have a complex wired network as well as wireless. Depending on the level and amount of data and applications running at the campus, we may or may not want to store the packets locally using an appliance with sufficient disk space, such as the Omnipliance Core. Or we may just want to monitor that area knowing that those packets are also funneling back through the NOC and Data Center, where we can archive them. We'll also need to look at capturing wireless if we have wireless users and look at what's going on at the users' actual desktops.

For the sample topology shown in Figure 6, we'd want at a minimum visibility into each of the core areas: the NOC, the Data Center, our Corporate Campus, and Remote Office. We'd also want to have storage available in the NOC and Data Center to archive packets to avoid having to reproduce network problems.

Management of the Data

Once your monitoring points are placed, you'll need to manage that distributed network monitoring data. Returning to the sample topology shown in Figure 6, we have multiple appliances at the NOC, Data Center, our Corporate Campus, and Remote Office. We also need to remember that some of the traffic is 10G traffic. In many locations, we're capturing to disk because we need to save all the packets.

If you want to conduct root-cause analysis, you need to be sure you're archiving the data for a long enough period of time such that the data will still be stored when you begin your analysis. At a minimum, you should calculate how much data is captured overnight or on a weekend. The amount of data collected off hours will typically be lower than during business hours.

Here are some guidelines for hard disk requirements, assuming steady-state traffic of 1 Gbps:

- 7.68 GB/min
- 460 GB/hr
- 11 TB/day

In addition to having the data available when you need it, you also need to have enough RAM to perform your analysis. Pre-compute the maximum RAM your search will use, for example, 10 – 128MB files searched with no limits (remember that by default all analysis options are typically on) could use 1.3GB. Turn off analysis options that aren't needed. For example, if you're not interested in VoIP or Video, you can turn off the Voice & Video Analysis option. The more analysis options enabled, the longer your search will take.

High volumes of data can also be addressed by using filters and slicing the data into manageable segments. Once you know what you're looking for you can exclude normal data. If you have a reoccurring evening problem that involves a particular server, you can turn on filters to capture packets just from that server and create a more manageable data set.

WildPackets Distributed Network Analysis Solutions

In order to make sure your applications are performing properly, you need to first determine what optimal performance is, and have tools in place that can perform 24/7 monitoring on your applications. With WildPackets

Beyond Monitoring – Root-Cause Analysis

Distributed Network Analysis Solutions you can monitor application response time, round-trip network delay, server responsiveness, database transactions per second, and a myriad of other low-level statistics. You can also use the Application Performance Index to standardize the overall assessment of user satisfaction.

A key factor in solving application performance issues is having the ability to analyze down to the packet level. However, having only a packet-level view is not enough. You need a distributed network analysis solution that provides both high-level monitoring capabilities that can keep you aware of how your applications are functioning from a business perspective, and that performs deep packet analysis when a problem does occur.

WatchPoint network monitor and OmniPeek Enterprise network analyzer, in conjunction with TimeLine network recorders, deliver high-level monitoring capabilities with one-click packet drill down while supporting both local and remote data capture and analysis—the perfect solution for distributed network analysis.

24x7 access to ALL network data lets you:

- Access and analyze real-time data vs. recreating issues post-case
- Easily share data, bringing problems to resolution more quickly
- Reduce and can even eliminate the need for “fly and fix” network engineers every time a problem occurs at a remote office

With WildPackets Distributed Network Analysis Solutions, you’ll have vital information for managing and troubleshooting your entire network at your fingertips 24x7.

Learning More

- **“Beyond Monitoring – Root-Cause Analysis”** With the introduction of NetFlow and similar flow-based technologies, solutions based on flow-based data have become the most popular methods of network monitoring. While effective, flow-based network monitoring has a significant limitation – it lacks the depth of data to perform true root-cause analysis. This paper introduces a new concept – Network Monitoring for Analysis, a technique that lets you know when there’s a problem and why there’s a problem, and illustrates a solution that provides the uncompromising performance and capabilities needed to address Network Monitoring for Analysis.
- **“Ending the Blame Game: Troubleshooting Distributed Application Performance”** Gone is the luxury of sending network engineers to physically visit a site to troubleshoot performance issues. Today’s geographically distributed companies and distributed applications require a 24x7 proactive approach to measuring and monitoring application performance. This paper defines Application Response Time (ART), presents factors to consider when troubleshooting the performance of distributed applications, and identifies must have criteria for distributed network analysis solutions.
- **“Network Forensics in a 10G World”** With highly utilized networks, capturing network traffic with individual SPAN ports and taps typically results in spotty overall visibility of your network. In today’s 10G world, you need a purpose-built network forensic solution in place capturing ALL network data, 24x7, to ensure a stable and safe network. This white paper identifies the unique challenges of highly-utilized 10G networks, establishes

guidelines for ongoing network data collection, and addresses the conflicting demands of traditional (TCP/IP) data analysis and VoIP analysis.

All of these white papers and more can be found at www.wildpackets.com under the “Resources” section.

About WildPackets, Inc.

WildPackets develops hardware and software solutions that drive network performance, enabling organizations of all sizes to analyze, troubleshoot, optimize, and secure their wired and wireless networks. WildPackets products are sold in over 60 countries and deployed in all industrial sectors. Customers include Boeing, Chrysler, Motorola, Nationwide, and over 80 percent of the Fortune 1000. WildPackets is a Cisco Technical Development Partner (CTDP).

To learn more about WildPackets solutions, please visit www.wildpackets.com, or contact WildPackets Sales: sales@wildpackets.com or (925) 937-3200.

Conclusion

Network Monitoring for Analysis is significantly different from Network Monitoring. Though the choices for Network Monitoring are diverse, you must keep in mind the compromises in data granularity, data accuracy, cost, and overhead that your choices imply. Network Monitoring solutions based on SNMP and flow-based technology lack the ability to perform true, root-cause analysis, meaning additional solutions, at additional expense, will need to be deployed to achieve true Network Monitoring for Analysis.

A core capability of a solution designed for Network Monitoring for Analysis is the capability of recording network data for post-incident analysis. Data recording eliminates the need, and the time and expense, of having to reproduce network anomalies – you will already have the data you need. Determining your analysis needs is critical both for specifying the appliances needed to adequately cover your Network Monitoring for Analysis needs, as well as optimizing the performance of this equipment once installed.

WildPackets offers a complete, fully-integrated solution for Network Monitoring for Analysis. By covering both your network monitoring AND your root-cause analysis needs, WildPackets provides the most cost-effective and easiest to use solution available today.