SANS ANALYST PROGRAM

Fourth Annual SANS 2008 Log Management Market Report

Demanding More from Log Management Systems

A SANS Whitepaper – June 2008

Written by: Jerry Shenk

Why Does Log Data Matter?

Why Are People Collecting Log Data?

How Are Organizations Using Log Data?

What Are Companies Using for Log Management?

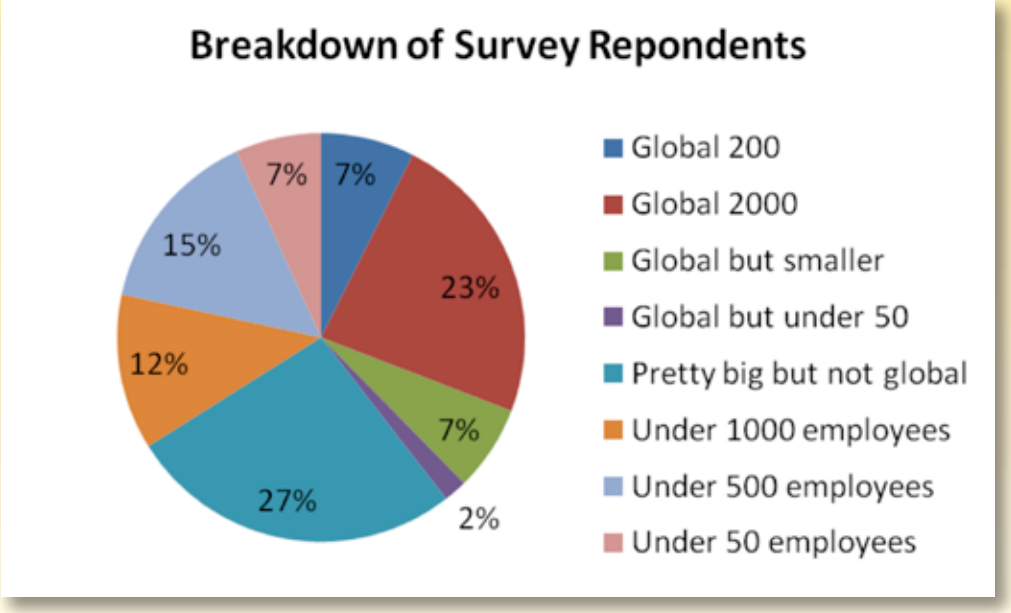
What Are the Pain Points with Log Analysis?

Executive Summary

The SANS Industry Analyst team conducted its first major survey on the Log Management Industry in the spring of 2005. Since then, year-after-year results have shown a growth in the number of companies that are maintaining and using log data.

According to this survey, these companies now want more data collection, correlation and analysis, but they work under ongoing limitations in these areas. As in past years, the lack of common format between log systems among (commercial and custom applications) is still causing problems with collecting, sorting and parsing of log data. This lack of interoperability is denying organizations access to data that they now know is in their logs – data that they can't access under existing conditions.

This, in turn, causes more dissatisfaction even as log management features and capabilities continue to improve.



Of those using their log data, they are using it most for detecting and analyzing security and performance incidents, and for minimizing downtime. Compliance reporting was also a big driver for collecting log data – up from 43 percent last year to 48 percent this year. The forensic use of log data ranked last in importance, which seems to conflict with the fact that organizations are storing data longer. And, while compliance is indeed driving more log storage and for longer periods of time, most are not using it proactively for compliance, they're using log data "after the fact." Clearly, they are having a difficult time realizing the full value of their log data, which is also supported by responses to other survey questions.

From 2005-2007, respondents grew increasingly satisfied with their log management situations. This year, that number slipped slightly, which is another indication organizations want more from their log data than they are getting. In 2005, 25 percent of those surveyed were satisfied with their logfile analysis. In 2006, that number rose to 28 percent and in 2007, it rose again to 37 percent. This year, that number slipped to 35 percent. There are many possible reasons for this decrease. Comments at the end of the survey include inadequate automation and correlation as top reasons for dissatisfaction, followed closely by lack of budget.

The one thread that shows up numerous places throughout this survey is a need for simpler processes of mining logs for valuable data that could point to weaknesses, intrusions, violations and inefficiencies that need attention and repair. This is difficult to achieve because there is little consistency among firewalls, IDS, routers and switches, operating systems, databases, production applications and myriad other devices and software across the enterprise that produce and store log data.

Why Does Log Data Matter?

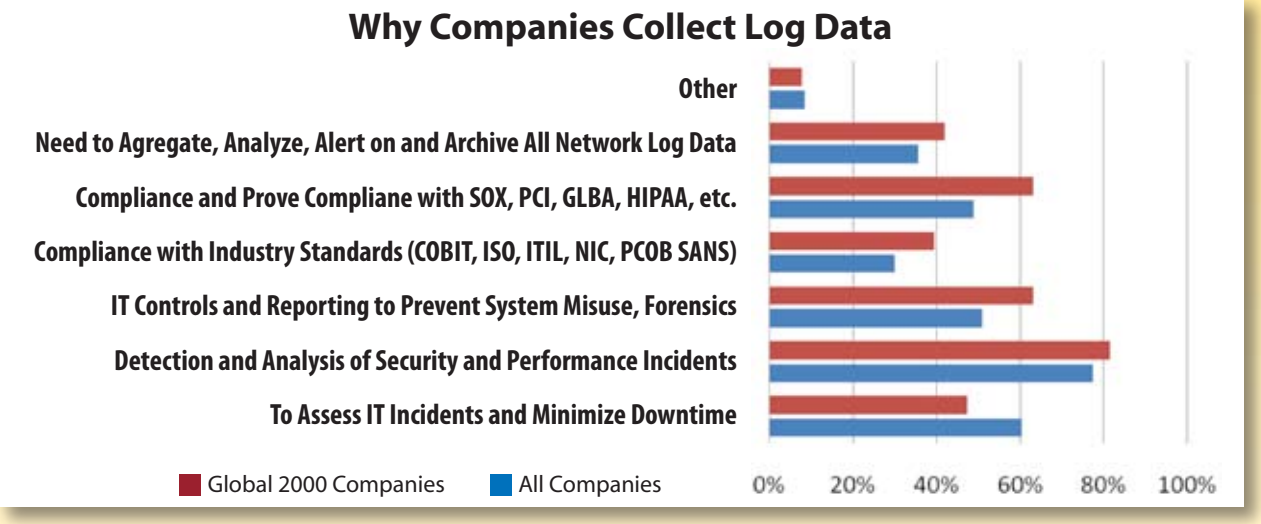
So, why does SANS conduct this Log Management Industry Survey every year anyway? Logs contain important information that can make the difference in passing an assessment, stopping an intrusion and knowing where to make repairs on the network. When something goes wrong, logs are always the first thing support technicians want to look at. That's why every good security course talks about turning on logging and gives ideas on how to tune logging.

Logs are how we tell what computers have been doing, something you can't see just by looking. Last summer, when I returned home from vacation to find water dripping from the kitchen ceiling, I didn't need any logs to know what was happening. I could see the water, water flows downhill, and I knew the bathroom was on the floor above it. We simply can't do that with computers. Log data is our only way to see what's going on and what went wrong. They can even sometimes give an early warning.

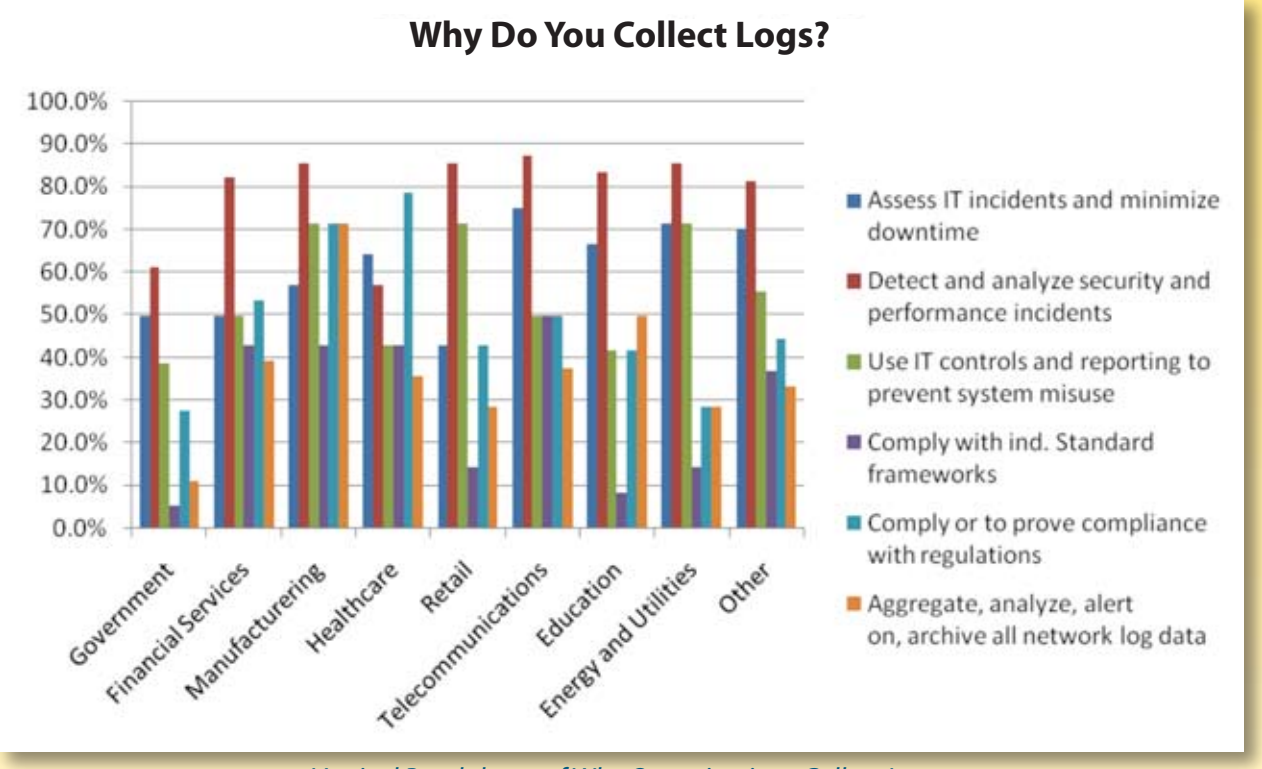
Log data matters because it gives us a way to see what our computers have been up to – using the term 'computers' very loosely. What we're really talking about is any device associated with computing – including switches, routers, firewalls, applications, databases, appliances and computers. The list of sources for logs is almost endless.

Why Are People Collecting Log Data?

In a question about how log management would most benefit their organization, 51 percent of respondents picked event detection. Trailing as the second most important option was day-to-day IT operations at 13 percent. Among the Global 2000, the same two options were first and second with 39 percent of that group picking event detection and 23 percent picking operations. Regulatory compliance was ranked third most important in both groups.



When these same questions are analyzed against the various industries represented by the survey, all segments – except for healthcare – stated that they collected logs because of their value to detect and analyze security and performance incidents. Not surprisingly, respondents from the heavily-regulated healthcare industry selected compliance with regulations or proving compliance with regulations and standards as their primary reason for log data collection. Remaining industries, including financial, ranked compliance third in order of importance. Among healthcare respondents, assessing IT incidents and minimizing downtime was selected second; detecting and analyzing security and performance incidents was third, and detecting and analyzing security and performance incidents was the fourth most important reason.



Vertical Breakdown of Why Organizations Collect Logs

How Are Organizations Using Log Data?

When determining how they most benefit from use of their log data, compliance came in fourth on the list. Respondents were asked to pick the three most important ways log management could benefit their organizations and rank them in order. This included security alerting, compliance reporting, forensics, system maintenance and information asset protection.

A rating average was derived from the choices that were made. "Security alerting" topped the list with a rating average of 2.17. Information asset protection followed closely with a rating of 2.15. System maintenance came in third with a rating of 2.08. The bottom of the list included compliance reporting with a rating of 2.00 and forensics with a rating of 1.34. What this means is that organizations who are strong in compliance also see other ways they can use this data to their benefit.

Analysis

In the 2008 SANS Log Management Survey, 78 percent of respondents said their reason for collecting log data was "Detection and Analysis of Security and Performance Incidents." That's up from 46 percent in 2006. In this year's survey, we dropped the word "automatic" from this choice, so it's safe to assume that part of the increase this year is due to the change in the wording. In itself, this is interesting because it reveals that some people are doing automatic detection and analysis of security and performance incidents, while others are also using logs for detection and analysis, but they don't have the automation piece.

However, the survey was also able to show that more respondents are doing automation this year than last year. The use of log analysis appliances is up from 10 percent in 2007 to 19 percent in the 2008 survey. When asked what they use automation for, the clear winner at 65 percent was "Use the logs after the fact to help troubleshoot." Some 25 percent indicated a homegrown application with daily automated review using keyword detection, up from five percent in the 2007 survey.

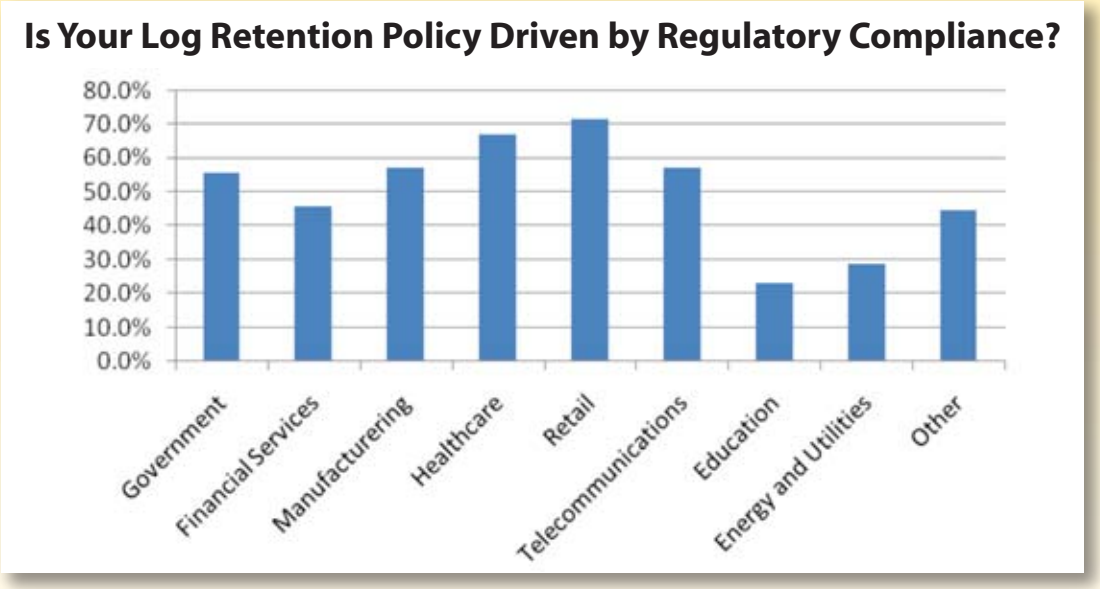
When it came to stating what changes they would like to make to their log management systems, one of the main responses was about the need for automation of alerting for system and security events. Another key change people would like is the correlation of events between different devices. What's telling is that, while automated appliance usage is up, correlation is slipping. In the 2008 survey, 32 percent do automated event correlation, as compared with 42 percent last year. Among the Global 2000 companies, 52 percent are doing automated correlation this year compared to 72 percent last year. Another area where organizations need more automation is in analyzing security and performance issues together, as we will explore later in this paper.

Log Retention

In another question, 47 percent of survey respondents say compliance is driving log retention policy at their organizations. In the 2008 survey, when people were asked how long they maintain logs, the largest group (19 percent) maintained logs for one to two years. In the 2007 survey, the largest group at 14 percent was "Operating System Default/Not Sure." In this year's survey, 35 percent of respondents indicate that they intend to keep log data longer than one year, up from 20 percent in the 2007 survey.

The majority of the respondents who selected the option "other" as their reason for collecting logs mentioned specific regulations or requirements and security issues. This amounted to 5 percent of respondents that have security- and regulatory-related reasons for log data collection but didn't feel that they quite fit into the categories that we had defined.

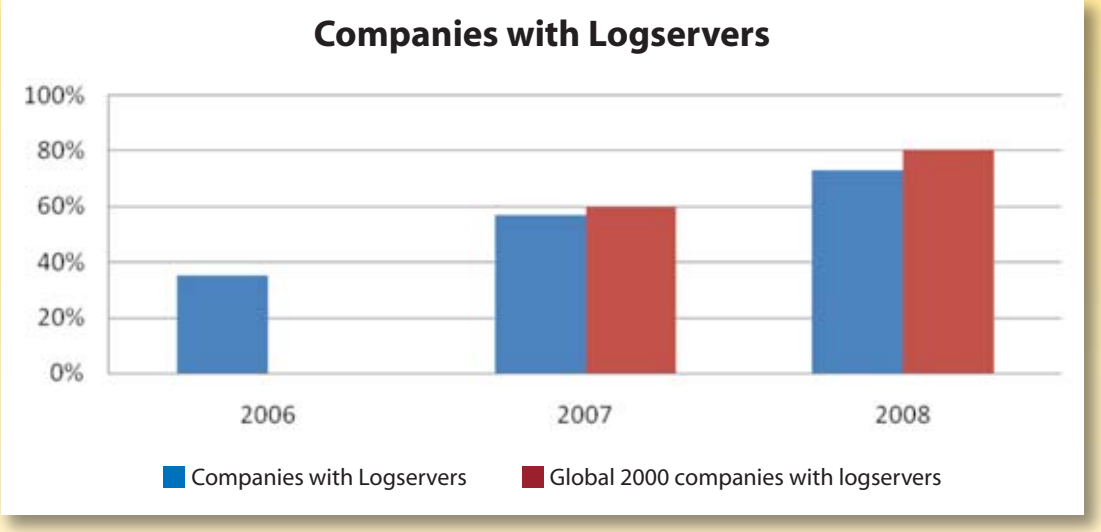
Compliance is a larger factor for some business segments than others. The retail segment seems to be the most affected by compliance issues. Some 71 percent of respondents in the retail segment, under new and evolving PCI DSS standards, said that their log retention policy was driven by compliance. PCI compliance has specific recommendations about the storage of log data. Some 66 percent of healthcare respondents also ranked regulatory drivers as their reason for retaining log data. Following that, 55 to 57 percent of respondents in the government, manufacturing and telecommunications sectors responded similarly.



Larger companies (Global 2000) collect logs for many of the same reasons, with a higher percentage of Global 2000 organizations collecting log data for compliance and regulatory purposes. The one category that the Global 2000 chose less often this year than overall respondents chose is the assessment of IT incidents and the minimization of downtime, which indicates log management is not being used to its potential in troubleshooting as well as alerting in larger organizations. This is down from the 2007 survey, in which both the Global 2000 and the other companies showed nearly the same amount of interest in this category.

What Are Companies Using for Log Management?

Seventy-three percent of respondents indicated that they had logservers in this year's survey. In the 2007 survey, 57 percent of respondents indicated that they had logservers – up from 35 percent in 2006. That is a particularly interesting statistic given that 64 percent of respondents are not satisfied with their log management solutions. Still, they are attempting some form of log management anyway, which indicates that they're aware of the value of their log data.

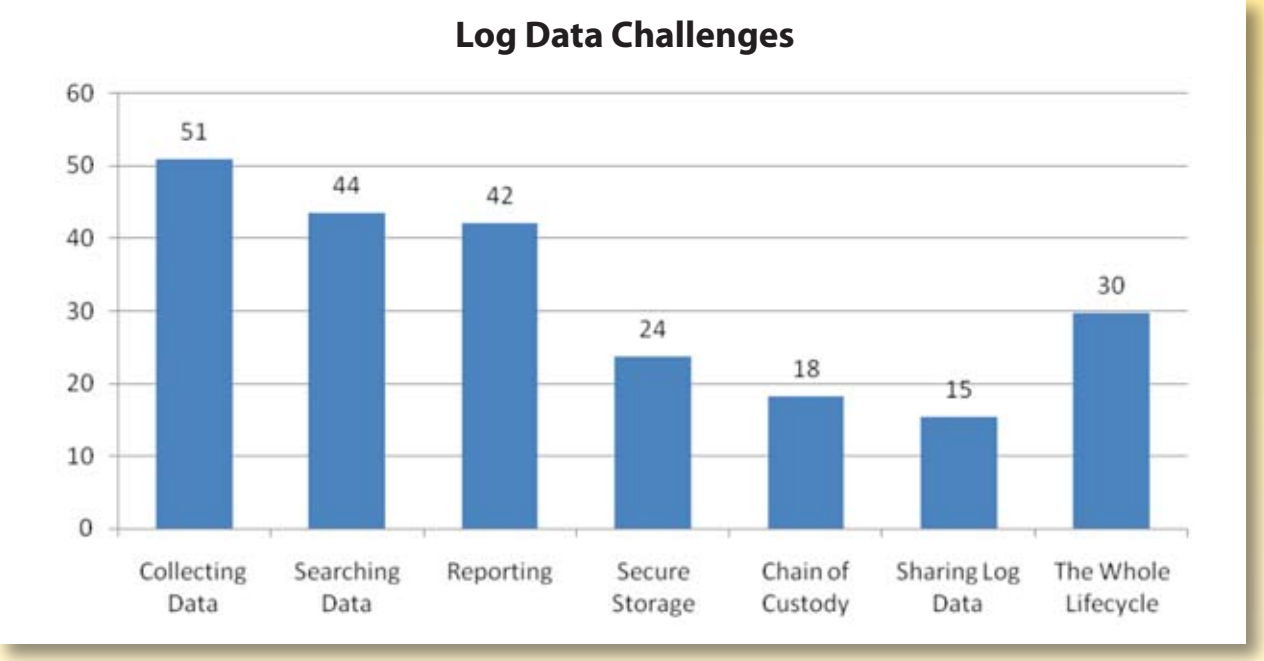


In this year's survey, instead of allowing for only one answer to what vendor is used for log management, we allowed respondents to check all that applied. This yielded a surprise – many companies use more than one log management vendor. This doesn't help with a comparison of log management vendors from prior years – some went up and some went down – but the "vendor" that was selected the most often was "Homegrown Solution." Thirty-eight percent of overall respondents selected this category as one of their log management vendors while 46 percent of the Global 2000 companies selected that option. It will be interesting to follow this number in the coming years.

We also asked them to rate their level of satisfaction. Thirty five percent of the respondents overall were satisfied with their log management, and 42 percent of the Global 2000 were satisfied. Of those companies that are satisfied, we examined what they were using. "Homegrown" was the largest single vendor category and 27 percent of both the Global 2000 and the total group of respondents were satisfied with their home grown management systems. Large vendor products tended to score less – between 17-25 percent overall satisfaction, with a few specialty vendors ranking 50 and one ranking 70 percent satisfaction.

What Are the Pain Points with Log Analysis?

In this year's survey, we asked people to rate aspects of the log lifecycle as critical, important and least important. Slightly over 50 percent of respondents rated collecting logs as their most critical issue. This was followed by searching log data and reporting on log data. Under the least important rating was sharing data followed by maintaining chain of custody.



The overall story from all of the responses is that getting useful data is too difficult but that it would be valuable.

In a "free-form" question at the end of the survey, people were asked for additional comments. Alerting, querying, reporting, correlating, and analyzing were the most common topics. A few respondents commented that "the right people aren't looking at the data," that "nobody is looking at the data," or that "people aren't making the necessary effort to understand the log data." One respondent commented on the difficulty of finding time in his day to implement a product that had been purchased. Another comment was simply, "start." Both of these comments point to a shortage of manpower. Lack of budget also was mentioned a number of times.

Collecting Logs

Just slightly over half (51 percent) of survey respondents ranked collecting logs as their most critical challenge in the log management lifecycle. This is actually an optimistic sign because it suggests that people are at least trying to collect and store their logs so they can get value from their log data. According to this year's survey, 80 percent are collecting, and 67 percent are archiving logs. In the 2005 survey, only two percent of companies surveyed stored their logs longer than one year. In this year's survey, 35 percent of companies stored their logs for a year or longer.

But as the survey suggests, there are still a number of issues related to the collection of log data holding organizations back from realizing the benefits of log analysis. The sheer amount of information needing to be sorted, organized and searched is overwhelming, and tools aren't simplifying this process well enough. In the first place, there's the problem of how to collect the data from the various clients and programs into the log management system. While this is being done a number of different ways, the two primary options are to have the log generating device send data to a log server or have the log server periodically pick the data up. Some systems are doing this well, and logging is straightforward, while other applications present log data in ways that can't be collected and stored easily, and still other applications don't even have logs.

Syslog servers are the de facto standard for sending data to a log server. In this year's survey, 75 percent of respondents who had log servers indicated that they are running syslog servers. This is up slightly from 72 percent in 2007. The standard syslog protocol is not perfect but it is widely used. However, some applications and operating systems still fail to support syslog servers. Windows file servers are one example of a server that does not support syslog. Windows event logs can be picked up by a logserver using the Windows Management Instrumentation Command-line (wmic).

One of the issues with syslog is that it uses UDP for communication. Since UDP is a connectionless protocol, it is relatively simple to spoof syslog messages. Another issue is that by default, encrypting syslog traffic is not supported. One possible alternative is syslog-ng which is an upgrade to the standard syslog that addresses these issues. Some sites will be able to make that option work for them. RFC 3164[1] defines the syslog standard. Syslog-ng is more complicated to set up than the older syslog protocol, and support from hardware and software manufactures is not as widespread.

[1] http://www.ietf.org/rfc/rfc3164.txt

SNMP traps are also used to send data to log servers. An SNMP trap is a message about an event that a switch, router or other device would send out. Many syslog servers will also accept SNMP traps. SNMP traps have the issues of spoofing messages and messages being unencrypted. RFC 1157[2] and RFC 1215[3] define the SNMP standard. Newer versions of the SNMP protocol make some effort to resolve these issues but at the expense of simplicity and wide support.

Some log servers use **ftp or file shares** to retrieve log data. There are also agents that can run on a Windows server to forward logs to a logserver. Some of these agents are proprietary to the logserver and some are standards-based and send the event log data to a syslog server. One example of a standards-based logging agent that would run on Windows server is the publicly available Snare agent[4].

Command-line search through Event Logs

Here is a simple command-line that will retrieve the NT Event log information from a Windows 2003 server:

```
wmic /node:192.168.1.195 /user:administrator
/password:password ntevent where (message like "%logon%")
```

In this example, the word logon is a variable that's being searched for. It might be interesting to search for "fail" or "cmd" or perhaps a username that should not be logging on. Searching through the event logs with a command-line utility can be more powerful than the graphical event viewer that is included

CAUTION: This can run for a long time before it returns results (could be 30 minutes or longer depending on size of logs). This command can also impact network traffic CPU utilization on the requesting computer. Test in lab environment before attempting it on production equipment.

Analyzing Logs

After collecting the data, searching (44 percent) and reporting (42 percent) are the most problematic areas of log management, based on responses. Complaints about both of these options were mentioned in an open-text comment option at the end of the survey, in which respondents said they would like to see improvements in alerting, querying, reporting, and correlating to alleviate problems associated with analyzing their disparate sources of log data. This could be a lifecycle issue, with more mature groups having achieved log data collection to a satisfactory level (49 percent), and the rest still trying to collect log data in the first place.

If software logs can start agreeing to say things the same way, that will make log analysis much more effective, which in turn should enhance security and system management. In last year's log management report, we noted that Mitre is standardizing the way log data events are expressed through a Common Event Expression (CEE) – a standard log language for event interoperability.[5] The syntax, transport and taxonomy specifications are underway, as Mitre works with the Open Group's Distributed Audit Service (XDAS)[6] to deliver its first taxonomy in the August-September 08, timeframe.

[2] http://www.ietf.org/rfc/rfc1157.txt
[3] http://www.ietf.org/rfc/rfc1215.txt
[4] http://www.intersectalliance.com/projects/SnareWindows
[5] http://cee.mitre.org/cee.html
[6] http://openxdas.sourceforge.net/doxygen/html/main.html

While we wait for vendors to make changes to the way their applications, operating systems and appliances express their log events, the next best way to handle the data is through normalization done through the log management system. Many log servers have filters and processors that will normalize data from disparate systems so that events can be detected across routers, firewalls servers and applications with their individual ways of reporting an event. They by no means catch all logs at this time, since proprietary applications will always be problematic. But they correlate enough logs from firewall, IDS, operating systems and other pervasive applications to get a good snapshot of events in their entirety.

Sharing Log Data and Chain of Custody

Behind collecting, searching and reporting on log data, respondents also had issues with sharing log data and maintaining chain of custody information on log data. Twenty-nine percent of respondents stated that the whole lifecycle of log data was a critical problem, and 50 percent said that it was an important problem. Looking at the Global 2000 companies, the number is a little different. Among the Global 2000, 41 percent indicated that the whole lifecycle was a critical problem and 45 percent indicated that it was an important problem. Sharing log data and chain of custody issues will become more important as the log management industry matures and IT personnel begin getting more of the data that they need.

In the 2008 survey, 28 percent of respondents indicated that they provide log data to other departments or executives. This is down from 37 percent in 2007. According to survey feedback, what data is being shared is being provided to security personnel and management. Only five percent of respondents indicated user activity monitoring is of top importance.

In one case, the information was being used for billing – an interesting notion when you consider the IT department is often seen as a cost center – and functions within IT like log management are well-suited for supporting such applications. If IT can develop a business support mindset, it may be possible to identify ways for log management to provide a financial return to the company. One familiar example is in monitoring web traffic. If visitors to a web site can be identified geographically, it may indicate an interest that could be targeted for development. Application licensing is another profit-driven use of log data that was mentioned by administrators.

Storage and Encryption

This year SANS posed a new question asking if centralized log management should be able to store log data encrypted. The answer was a resounding yes, with 51 percent in favor, 29 percent against, and the remainder undecided. This is in keeping with industry events making full disk encryption quicker and easier with minimal performance impact, and the ability of most commercial log management systems to store large amounts of data through a variety of means.

Summary

The largest single change in the 2008 SANS Log Management survey was how the Global 2000 market responded to the question of how much they spent on logfile analysis. In the 2007 survey, 59% of respondents chose "No clue, we look at logs as needed or when there is an issue." In 2008, only 14 percent chose that option. This is a strong indication of a growing awareness of the need for log management. The largest group indicated that they spent under $25,000 annually on log analysis. The average spending on logfile analysis by the Global 2000 works out to $190,000 each, or a $380 million annual market. It is a market holding steady with 2007, in which we estimated the market at $374 million.

While survey responses clearly indicate that organizations are becoming more aware of the value of log data in security, compliance and maintenance operations, they still have a long way to go. Sixty-two percent of overall respondents say that ROI is not used as a measure. That means that the vast majority are not even looking for ROI, even though in one case logging was used for billing, a cost center activity.

Organizations are looking for ways to increase their visibility into their log data, putting demands on tools vendors to continually improve integration, analysis and storage capabilities. In the meantime, many companies are tackling the log management problem with a combination of homegrown and commercial tools as they begin to make use of the valuable data lingering in applications and devices across their organizations.

About the Author

Jerry Shenk currently serves as Senior Analyst for the SANS Institute, and is the Senior Security Analyst for D&E Communications. Since 1984, he has consulted with companies and a variety of financial and educational institutions on issues of network design, security, forensic analysis and penetration testing. His experience spans small home-office systems to global networks. Along with some vendor-specific certifications and a CISSP certification, Jerry holds five GIAC GOLD certifications: GCIA, GCIH, GCFW, GSNA and GCFA – all completed with honors.

SANS would like to thank this paper's sponsor: