

Sponsored by
ArcSight, LogLogic, NetForensics, Novell, RSA and Trustwave

SANS Sixth Annual Log Management Survey Report

A SANS Whitepaper – April 2010

Written by Jerry Shenk

Survey Sample

Collection and Uses of Log Data

What Logs Are Being Collected?

Searching and Reporting

Log Data Storage

Tips for Managing Log Data

Advisors:

Dave Shackelford, SANS analyst and instructor,
director at Sword & Shield Enterprise Security

Barbara Filkins, SANS analyst, GIAC Gold, GCIH
(Silver) with specialty in HIPAA privacy and security

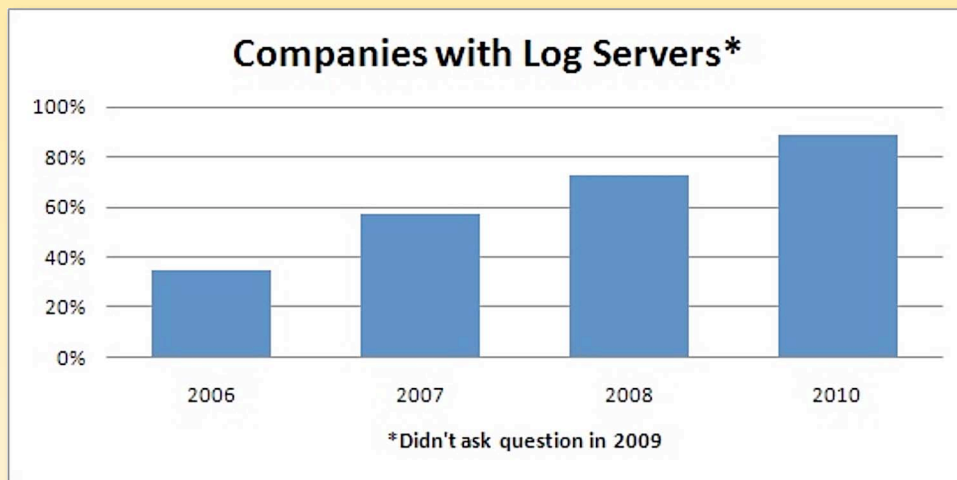




Executive Summary

Every spring since 2005, the SANS analyst team has released results of its annual log management survey. These surveys have tracked the progress being made in the log management industry and in the organizations using log management.

Year after year, more organizations say they have log servers to support their log management objectives—89 percent in this year’s survey, compared to 43 percent in 2005. Our 2008 survey built a case for the value of collecting and monitoring logs. This year’s survey reveals that this trend continues to expand: Companies better understand the importance of collecting logs for a growing number of purposes and are trying to derive more value from the logs they’re collecting. In this year’s survey, respondents have found improvements in detection, forensic and compliance uses of their data.



However, responses show there is room for improvement in the ability of log management systems to deliver value from logs being collected, specifically in the areas of searching (where 36 percent of respondents reported problems), and analysis (where 34 percent had problems). Whereas in 2005, respondents reported their biggest problem was simply collecting log data, in 2010, collection was cited by only 10 percent as the biggest problem—and 27 percent said that collecting log data was their least challenging problem.



Given the pervasiveness of Windows operating systems throughout the industry (and with their notoriously collection and analysis difficulties), a question was added to the survey this year on Windows log collection. Less than half of survey respondents were satisfied with their Windows log management capabilities and tools, indicating they collect the data, but analysis and reporting on the data are problematic.

In keeping with the drive to collect more logs and derive more value from them, this year's survey reveals that organizations are branching out from the standard sources for log data (servers, firewalls and routers) and are also collecting logs from in-house applications, HVAC systems and other physical plant sources. However, responses also show companies are having a hard time wading through and making use of the log data that they have collected. Even still, they find the data they do analyze to be useful for a growing number of purposes—particularly in the areas of detecting and preventing unauthorized access and of meeting regulatory requirements.

One area in which respondents would like to see improvement is the use of log data to enhance day-to-day operations. Quick access to logs can reduce the cost of resolving problems and help support other operational needs. Survey responses indicate some awareness that logs could also be used as a business cost enhancement, for example for chargeback billing and other purposes.



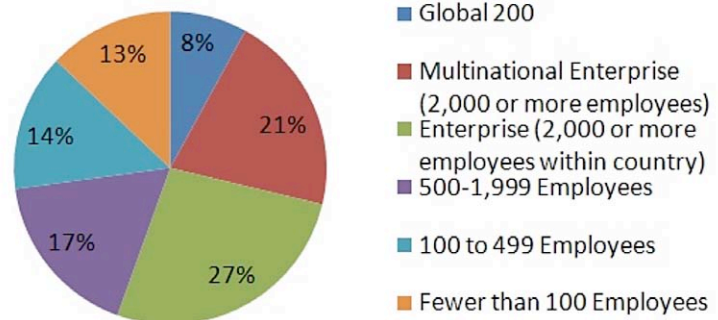
Survey Sample

Organizations of all sizes responded to this year's survey. Of the 501 who answered the question about organizational size, 27 percent were from enterprises of over 2,000 employees within a single country), 29 percent were from enterprises of 2,000 or more employees in multiple countries (this includes the Global 200), and 17 percent were from enterprises with 500 to 1,999 employees. The remaining 27 percent represented organizations with fewer than 500 employees. The vast majority of respondents held staff positions (rather than being consultants), with an equal mix being senior level and hands-on IT staff.

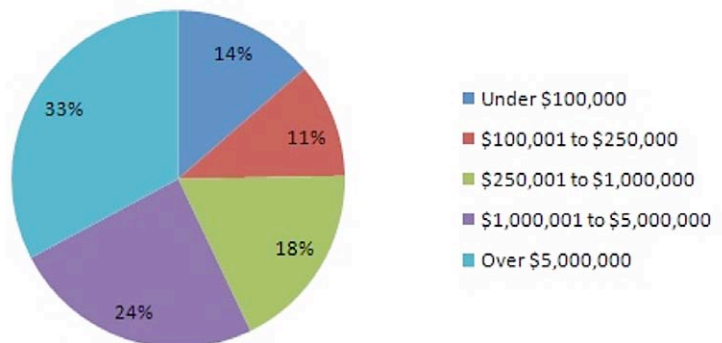
The largest group of respondents, 33 percent, represented organizations with overall IT budgets of \$5 million or more, followed by organizations with budgets of \$1 to \$5 million (24 percent). The remaining 43 percent had budgets of less than a million dollars.

A wide variety of industries were also represented: 19 percent from government, 17 percent from financial organizations, 11 percent from education, nine percent each from healthcare/pharmaceuticals and telecommunications, and the remainder from retail, manufacturing, utilities and engineering/construction. Throughout the survey, the issues reported are, in large part, consistent no matter the size of the company.

Breakdown of Survey Repondents



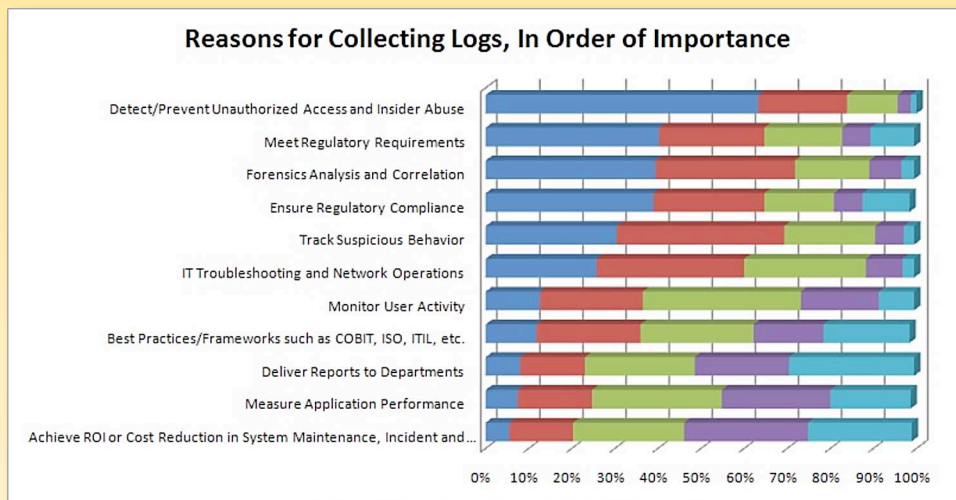
Overall IT Budget





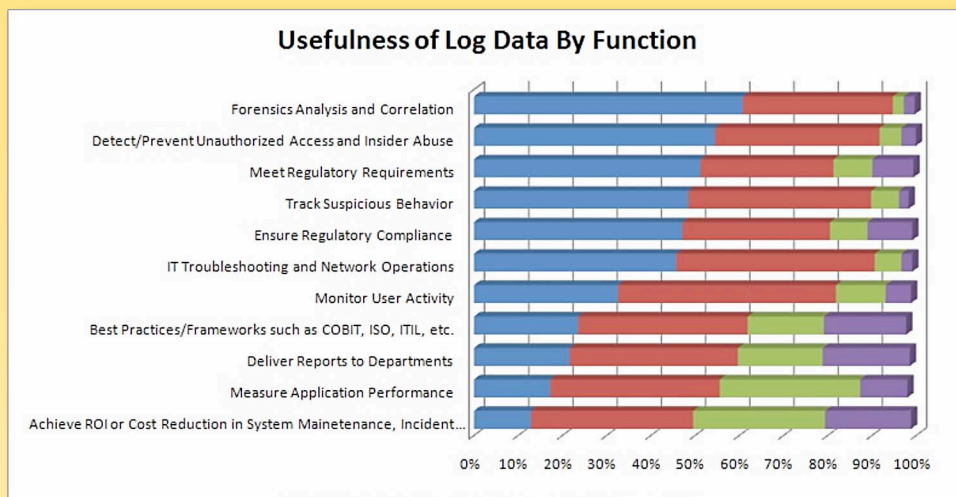
Collection and Uses of Log Data

In the 2010 survey, the top reason for collecting logs was to “Detect/prevent unauthorized access and insider abuse,” with 63 percent of respondents rating this as most critical. On a scale of one to five (one being highest), 83 percent rated detection/prevention in the top two rankings of criticality (1 and 2). Second place was a statistical tie between “Meet regulatory requirements” (41 percent), “Forensic analysis and correlation” (40 percent) and “Ensure regulatory compliance” (also 40 percent). Some other items that were picked as critical include “Track suspicious behavior” (31 percent) and “IT troubleshooting and network operations” (26 percent).



These reasons for collecting logs closely align with the usefulness of the data being collected, particularly in the area of monitoring. More than half of this year’s respondents found log data to be very useful in forensic analysis and correlation, detecting/preventing unauthorized access and insider abuse, and meeting regulatory requirements. In addition, logs are also being utilized to “Track suspicious behavior” by 49 percent of respondents.





In the 2009 survey, “Track suspicious behavior and user activity monitoring” was selected by 74 percent of respondents as a critical reason for collecting log data. In this year’s survey, 91 percent of respondents indicated that this was important or critical, and 91 percent indicated that log data was useful or very useful to track suspicious behavior.

In comparing the usage information with difficulties respondents had with the reporting and searching aspects of log management, it appears organizations are collecting logs, but still need to improve in order to derive all the value those logs have to offer. Respondents indicated they were able to get the data they needed to track suspicious behavior and network problems, but they had to work too hard to extract the data and make sense of it. Results point to immature searching and reporting capabilities as the primary reasons for this, as revealed in another question about the most challenging aspects of the log management lifecycle. In it, “Searching through data” and “Analysis and reports (and ability to interpret results)” ranked as the top challenges (36 percent and 32 percent, respectively).

For now, these organizations aren’t deriving the deeper value log data can offer. In the 2009 survey, the deliver reports to departments category was ranked as important by 23 percent and least important by 53 percent of respondents. This year, 29 percent of respondents consider delivering reports as critical to important, with 20 percent of them finding their log data as very useful and another 38 percent finding log data somewhat useful for these purposes. This indicates improvement in log data usefulness, but also indicates there is room for much more improvement.



Using logs for IT troubleshooting and network operations was ranked as critical for 26 percent of respondents, with 46 percent of respondents ranking it very useful and another 45 percent ranking it somewhat useful. In the 2009 survey, troubleshooting was ranked as critical by 14 percent of respondents, with 49 percent ranking it as important. This indicates that organizations aren't installing log management systems for their troubleshooting value, yet an increasing number are finding troubleshooting assistance to be a useful added bonus.

In another question about measuring their security effectiveness, respondents ranked meeting compliance targets as the most critical (41 percent). "Incident remediation time/success" was close behind, followed by "Measuring security state of systems/network." Regulatory requirements such as PCI DSS, HIPAA, GLBA, SOX, FISMA and others will continue to push the log management field for the foreseeable future. This was one of the areas that the 2005 SANS Log Management report predicted would be a future driver. Reducing overall security costs was also ranked as important by 50 percent of respondents.





What Logs Are Being Collected?

More organizations are also using log servers, with 41 percent using two to five log servers, 21 percent using a single log server, and 15 percent having more than 20 log servers. The remaining organizations have between six and 20 log servers. The biggest single difference between multinational enterprise companies (more than 2,000 employees) and single-country enterprises is the number of log servers (28 percent of multinational enterprises had more than 20 log servers, whereas only 11 percent of single-country enterprise organizations had more than 20 log servers). The multinational group also did more log collection from desktops. They also did more tracking of suspicious behavior and found that tracking to be more useful. Collecting logs from desktops and the ability to monitor users are some features that any widely dispersed company (even multiple sites within a single country) can use to help manage and support multiple sites.

In this year's survey, 95 percent of organizations collected log data from "Firewalls, routers, switches, IDS/IPS, etc." We're also seeing a large increase in the amount of log data being collected from a variety of other types of sources. This suggests that the log management process is gaining popularity as a standard part of IT. The largest group of respondents (40 percent) is collecting log data from less than 100 sources; 36 percent collect data from between 101 and 999 devices; and 16 percent collect data from over 1,000 sources. (The remaining eight percent didn't know how many sources logs were being collected from.)

After security and network devices, the next most prevalent source of log collection is servers and mainframes (90 percent). No surprise there—considering servers and mainframes are where organizations carry out their most critical data processes.

Applications account for the next largest source of log data being collected. Last year, 41 percent of respondents collected log data from homegrown applications. This year, the definition was expanded to include both homegrown and commercial applications, and 73 percent of respondents say they're collecting logs from these sources. Clearly application logs present a growth area in log management: the ability to allow custom applications to be integrated into the log management system without reliance on the vendor.



Another popular log source in this year's survey was databases, with 73 percent of respondents collecting logs from them. Databases have their own logging systems that have improved over time and provide logging from within the database itself. Log management systems should be able to work with native database logs as well as logs generated from database management products, access controls, and so on.

About 50 percent of respondents also collect logs from "Identity sources (directories, IAM, IDM)." This includes Microsoft's Active Directory, Novell's eDirectory (formerly Network Directory Services), LDAP (lightweight directory access protocol) and other systems that control access to network data. IAM (identity access management) and IDM (integrated data management) systems are conceptual terms relating to controlling users and their interactions with data on a network.

Logs from desktops are also being collected—49 percent of respondents report collecting data from desktops. We did not ask about desktops in the past, so we can't relate this to historical calculations. Recent trends in attack traffic indicate that workstations are a primary target of attackers. For example, in a penetration testing webcast for CORE Technologies (since removed), it was possible to track the exploitation of a workstation using data from the event logs of the workstation. By sending log data from the workstation to a log server, this exploit data is made available for forensic analysis. With 40 percent of respondents to this year's survey indicating that "Forensic analysis and correlation" is of critical importance for their organizations, end point device logging will be a critical management element.

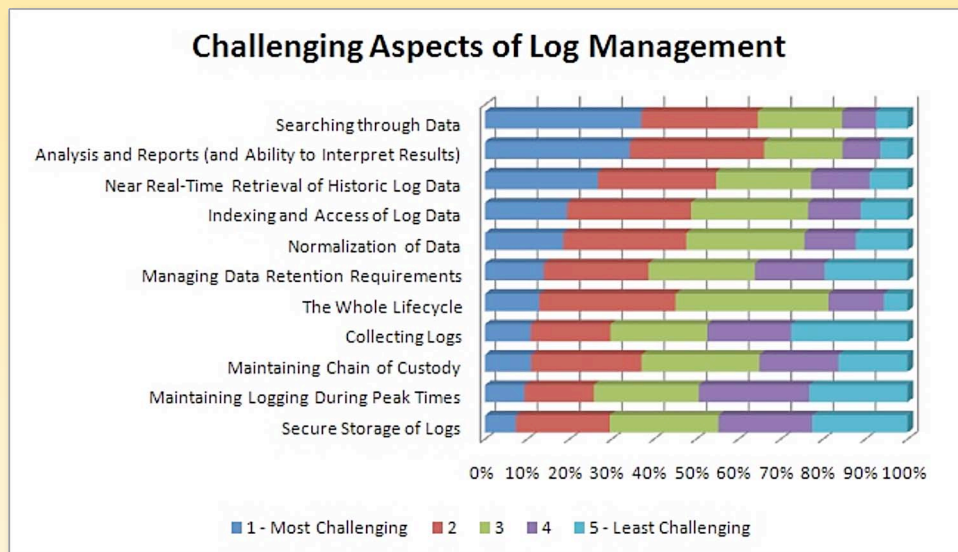
In the 2009 survey, there were a number of comments about organizations collecting log data from physical devices such as badge access, plant control systems, HVAC, and so on. This year, we specified these in our questions, and 48 percent of respondents indicated they collect log data from "Physical devices (badge access systems, plant control systems)." It will be interesting to see this use of log servers increase in the future as they become more integrated into operational control systems.



Searching and Reporting

Collection was the primary problem cited in our 2005 survey. Today there are a variety of solutions to meet the challenges of collecting log data from disparate sources, so collection is no longer a priority issue based on responses.

With the problems of collection resolved, the critical area of need is in analysis and reporting, which has been increasingly evident in the SANS Log Management reports for the past few years. This is the first year that analysis and reporting was picked as the most challenging aspect of log management by 64 percent of survey respondents who considered “Searching through data” their first or second most challenging aspect. “Analysis of reports and ability to interpret results” was the second most common problem area cited, with 65 percent selecting analysis as the first or second most challenging aspect in their log management programs.



Some of the key reports respondents desired include failed attempts to access accounts and resources—especially administrative access—and changes to users and groups. Reports that show deviations from normal activity patterns would also be nice, and providing access to such reports is one of many areas in which log management reporting capabilities are improving.

Log management vendors have been making an effort to include good log reporting options, but there are issues that make reporting difficult. Resources, including the SANS “Top 5 Essential Log Reports,¹” can help identify necessary reports. One of the biggest problems is that each software and hardware component (even some from the same vendor) will log data differently. Sometimes slight differences during an upgrade or change to the system will break existing reports because the log syntax was changed during the application upgrade.

¹ www.sans.org/security-resources/top5_logreports.pdf

This problem with differing log data formats might also explain why the majority of organizations (56 percent) are using more than one log management vendor solution. Duplication of solutions may be an attempt to work around some of the differences in data types by capitalizing on the strengths of various vendors' offerings. It is also possible that different groups within the organization chose multiple log management solutions to address their problems with reporting and analysis. Although access to multiple solutions may provide them with better reporting, it appears to reduce the efficiencies that using a single log management platform for centralized management could provide.

The use of multiple log management vendors to aid in reporting and analysis appears particularly true in the case of log management for Windows devices. In comparing the degrees of Windows log management success, the organizations using a single vendor's log management solution (including open source as a vendor) were most satisfied with the *collection* and *storage* aspects of their log management solutions, whereas organizations using two vendors were most happy with the *reporting* and *analysis* aspects of their solutions. This indicates that two solutions combined create better reporting and analysis capability.

However, using too many vendors may be detrimental to the usability of log reports. Organizations using three or more log management solutions were clearly the least satisfied, especially in the analysis and reporting areas. Minimizing the number of log management solutions an organization needs to use is an area in which log management vendors need to improve, supporting the major players in the industry so that consumers don't need multiple vendor products to manage their variety of logs. However, it's also an area in which different IT groups need to work together to increase efficiency and effectiveness.

Dissatisfaction with reporting and analysis indicates vendors are still having problems with log data normalization. *Normalization* is the process of converting the different ways various log sources report events by using common terminology. Industry efforts at developing common log data frameworks seem to be faltering. The best long-term solution is for the log management vendors to normalize log data and create critical, informative reports. Part of the responsibility falls on the log source vendors of the routers, firewalls, operating systems and applications that produce the logs necessary for system monitoring. These vendors need to maintain consistency within their product lines, minimize changes with each upgrade, and work toward some type of flexible, common logging language.

In summary, searching, reporting and analyzing are issues on which log management vendors are improving. Things are definitely getting better, but there is still room for improvement.





Log Data Storage

One of the trends we've noticed over the past few years with the SANS Log Management Survey has been an increase in log data retention time. In the 2005 survey, over half the respondents retained logs for 90 days or less. By 2008, 63 percent of respondents retained logs for one year or less. This year, the largest group retained logs for one to two years.

In many cases, the type of log data dictates the length of retention. This year we asked respondents how long they kept different types of logs. In general, the regulated data is stored longer than the non-regulated data, with PCI DSS, SOX and GLBA data being statistically tied for the longest storage time (48 percent of respondents stored SOX data for over one year and 49 percent of respondents stored GLBA data for over one year.) Respondents retained HIPAA and FISMA data for only a slightly shorter time frame. Thirty-two percent of respondents stored PCI DSS data for over one year; however 24 percent stored the PCI DSS data for 120 days to one year, as opposed to a slightly lower number for the SOX and GLBA respondents. Because PCI DSS places a one-year retention time frame on much applicable data, many of the PCI DSS respondents may have been deleting the data after one year.

One surprise was the number of people who stored regulated data for "Operating System Default" for "less than 90 days." This is particularly surprising with the amount of attention given to PCI DSS and other similar requirements recently. Across all types of data (including regulated data), roughly 25 percent of respondents were in this category. This indicates that either these organizations have no regulatory requirements to retain data, or they just aren't doing it.





Tips for Managing Log Data

The amount of log data being collected is growing at the rate of 15 to 20 percent per year, based on averaged estimates from respondents. Some companies report a 100 percent increase per year. The top factors that respondents expect to impact future growth are “Increased log sources” (47 percent), “New regulations” (32 percent), and “Inclusion of application logs” (27 percent). Regulatory drivers are consistent with current reasons for collecting logs, but these responses show organizations will continue to expand their uses for logs and to derive value from the log data.

The more logs collected, the more logs to manage—and it’s important for organizations to take ownership of the information in the logs they’re collecting. Organizations need to put a roadmap in place for what they would like to use their logs for and phase in more logs as they become useful.

Here are some tips for managing log data being collected:

- 1. Use Data Reduction Techniques.** The amount of available log data can be staggering. It isn’t difficult for a moderate-sized network with a router, firewall and a few servers to generate over a gigabyte of log data per day . You never know when data you collect may be useful, so it’s important to store all you can and eliminate what you don’t need. Some devices will allow for some granularity in what log data is stored. With a PIX or ASA firewall for example, it is possible to have the firewall logging data in debug mode but exclude a single error message with the command “no logging message xxxxxx” where xxxxxx is the number for a specific error message. For example, “no logging message 302016” will stop logging the “Teardown UDP connection” message on a Cisco ASA firewall.
- 2. Exclude Data From Your Searches.** Sometimes, a search turns up a LOT of hits—too many to be able to process. When this happens, people try to find the right search string to find just what they’re looking for. It sounds like an oxymoron, but to find useful information you start excluding information. For example, if you are looking for an outbound connection through your firewall, start eliminating things that you know aren’t a problem. If your web servers are hosted outside your network, exclude log data on those IP addresses. As you identify normal log events, start excluding them from the search results.

² www.sans.org/reading_room/analysts_program/eventMgt_Feb09.pdf



3. Know Your Data. People often look at log data only when there is a problem, which is a mistake because you need to know what data is normal in order to identify data that is not normal—and therefore related to the problem. Start looking at your logs today, and include asset owners in the log review process periodically. Don't go crazy—you can't read all your logs, but you can browse through them. Depending on what your options are for browsing through your logs, you may be able to exclude data to make your browsing more efficient. For example, if you are looking through Windows event logs, you will soon identify messages that are repeated quite frequently. If these messages are not indicating that there is a problem, eliminate them from your viewing screen.

4. Know Your Logs. A dramatic change in size can indicate a problem. Take into account the difference in days of the week. Typically, weekends have much smaller logs than weekdays. For some organizations and some specific servers, however, the opposite may be true. Get to learn what *normal* is for your organization. If your log management system has the capabilities, have it e-mail you a summary of log events every day. You'll soon learn what a normal summary for your company looks like. Correlate events you see in your logs with normal business activities, such as close of the business day, month, or quarter; the annual budget cycle; or specific audit events required in your industry.

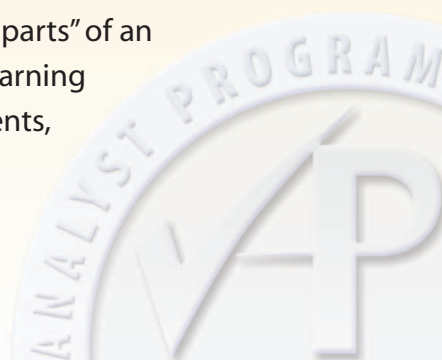
5. Research What You Don't Know. If you don't know what a log message means, look it up. Most of the log management tools make this easier, but most environments include log data that is not supported. Often, you can paste the "stock parts" of an event into a Google search window and get a good start on learning what the message means. If you are researching windows events, the site, EventId.net, is a helpful resource.

Watch What You Eliminate

Be careful when you eliminate data. Some "meaningless data" could be invaluable during the investigation of a forensic event. For example, on an ASA, a teardown messages will be logged every time an active session is disconnected. At first glance, this seems to be a lot of worthless overhead. These messages report the length of the connection and amount of data transferred in that session. This can be valuable information to have if you are trying to track down an internal data leak. If your log server is running on a variant of Linux, the following command can extract all the "Teardown TCP connection" messages that indicate the close of a session that transferred over 10 MB of data.

```
grep 302014 /var/log/messages | grep "bytes [0-9][0-9][0-9][0-9][0-9][0-9][0-9][0-9]"
```

There are more efficient ways to run this command, but this syntax is fairly readable. The command looks through the file `/var/log/messages` and uses `grep` (a pattern matching utility) to find the word "bytes," followed by a space character, followed by at least eight numbers. Each "[0-9]" represents one number, so you could add or subtract them to increase or decrease the size of the file transfer that will match. You can also change the first 0 to a 5 to only match on transfers 50 MB or larger.





Summary

Log management is a part of our computing infrastructures and the logs being collected and their uses are continuing to expand. Regulatory requirements continue to drive organizations to enhance their log management systems. More and more organizations are also finding log management systems to be useful for detecting and analyzing suspicious behavior and troubleshooting network issues.

Organizations are still having a hard time finding ways to get value out of their log data. Once again this year, analyzing data from different types of logging devices and generating reports is a critical problem for many organizations.

Survey respondents are finding growing uses for their logs, indicating that log management vendors are improving their analysis and reporting capabilities, as well as integration with other security and monitoring systems.

We expect to see that regulatory requirements continue to be a driving force in getting organizations started with log management. We also expect that organizations will continue to find uses for log data that tie back to their bottom line. Such uses include detecting and resolving network problems and detecting unauthorized access on desktops, laptops and mobile devices.

We also expect to see continued enhancements in reporting provided by log management vendors, by using tools that add intelligence to the analysis process through better correlation, presentation and decision support capabilities.

Log management is becoming more deeply embedded in day-to-day operation, as evidenced by survey respondents wanting management of logs coming off physical plant and other nontraditional log sources. The widespread collection of logs also indicates log management is at the beginning of its maturation process, with concerns now focusing on how much organizations can do with the data their management systems are collecting.





About the Author

Jerry Shenk currently serves as a senior analyst for the SANS Institute and is the senior security analyst for Windstream Communications in Ephrata, PA. Since 1984, he has consulted with companies and 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, Jerry holds six GIAC certifications, all completed with honors: GCIA, GCIH, GCFW, GSNA, GPEN and GCFA. Five of his certifications are GOLD certifications.



SANS would like to thank its sponsors:



The Security Division of EMC

