SANS ANALYST PROGRAM

Sponsored by
ArcSight, LogLogic, LogRhythm, Splunk and Trustwave

SANS Seventh Annual Log Management Survey Report

A SANS Whitepaper – April 2011

Written by Jerry Shenk

Survey Sample

Why Companies Collect Log Data

Users Want Better Log Data (and More of It!)

Top Challenges to Effective Log Management

Advisors:

Dave Shackleford and Barbara Filkins

Executive Summary

Every spring since 2005, the SANS Log Management survey has tracked the growth and maturity of the log management industry. This survey has consistently identified areas in which organizations are focusing their log management initiatives and continues to provide a roadmap to the industry for future improvement. Over the years, these surveys have shown growth in the collection and use of logs for security and compliance. Most recently, in the past two years, these surveys have shown that organizations are seeking more uses from their logs, but they have problems getting the value they want from those logs.

When this survey started seven years ago, log collection was only being done by 43 percent of respondents, compared with 89 percent who indicated they collected logs this year, which is consistent with last year's survey. So, log collection is no longer as much of a problem as it was in the past. Now, they're also collecting logs for much more than detecting suspicious behavior and troubleshooting, as in the recent past. Over the past two years, more respondents are also collecting logs for use in forensic analysis and correlation and to meet/prove regulatory compliance. In fact, these three uses for logs rank close enough in importance that it is fair to say that for a log management solution to be effective today, it must support all three.

In addition to the above top three uses, organizations are collecting more data from physical plant/operations systems (e.g., HVAC, SCADA), mobile platforms, and point-of-sale (PoS) devices. This means more log types to collect and analyze—each with their own data formats that can vary widely. Even when these log data format differences are slight (such as one date format being MMDDYYYY and another being MM-DD-YYYY), they must be adjusted in order to accurately correlate and report on the data. This has been an ongoing problem for users of log management technologies, particularly as they start to use their logs for more purposes.

In addition to normalization, respondents are also struggling with searching, correlating and reporting functionalities. Figure 1 illustrates the aspects of log management that respondents considered most challenging or moderately challenging.

Figure 1. Log Management Challenges

The mechanics of collecting, storing and archiving the log data are no longer the challenge in today's world of almost unlimited data storage. The challenge now is extracting the needed information for monitoring, management, compliance and decision-making (often in near real-time) from what respondents say is upwards of 100,000 events recorded per day.

This year, respondents were asked specifically about what was and was not useful in terms of searching and reporting capabilities. They selected real-time alerts as their most useful feature. However, they were less enthusiastic about their log management system's ability to interface with third-party tools or larger SIEM environments. Users also cited problems with correlation, searching and interfacing with heterogeneous systems, and difficulties locating information within logs.

In particular, Windows systems are still difficult to draw and normalize logs from. This is a primary problem for organizations this year, as in years past, according to responses. Windows, pervasive throughout most industries, is widely criticized for its unfriendliness to log analysis. However, all vendors of log management applications are making their systems interact better with multiple sources of log data, including from Windows systems. However, as one commenter wrote, all vendors still need to get better at generating useful events.

Despite shortcomings respondents report, organizations are increasingly dependent on log management to support core business functions including cost management, service level and line-of-business application monitoring, as well as more traditional IT- and security-focused activities, according to responses. The rest of this report details what organizations are doing with their logs today and what they still want from their logs in order to achieve the highest value for their business, security and compliance operations.

Survey Sample

A total of 747 organizations started this year's survey, with 571 completing the survey all the way through to the end. Organizations represented in this year's survey (see Figure 2) encompassed a wide range of industries and sizes. The largest industry verticals represented were financial (19 percent) and government (18 percent). Healthcare and education were well represented as well. The additional 23 percent that replied "other" included good representation from software companies, entertainment, managed services and consultants working among these verticals.

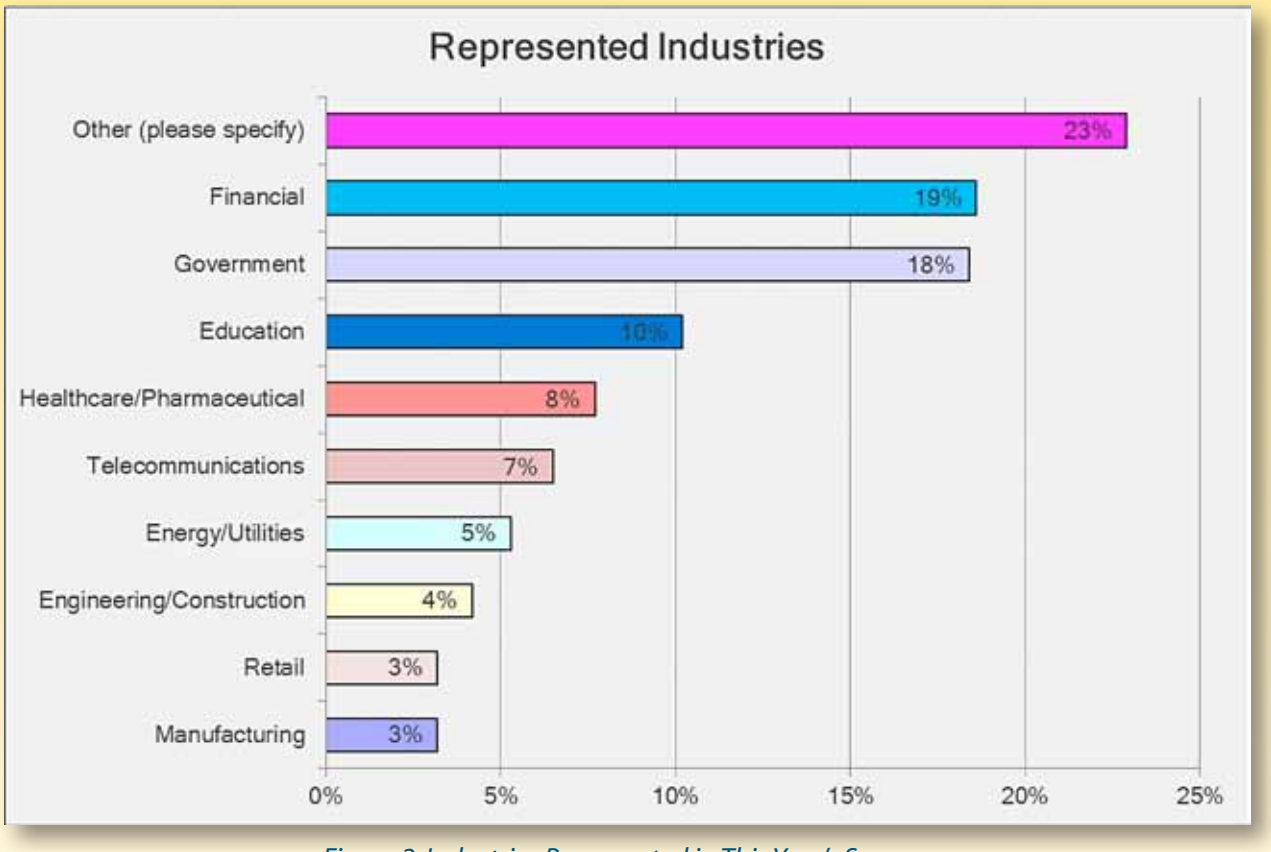

Figure 2. Industries Represented in This Year's Survey

Respondents were nearly equally balanced between large organizations (over 2000 employees) and mid-sized and small organizations, as shown in Figure 3.

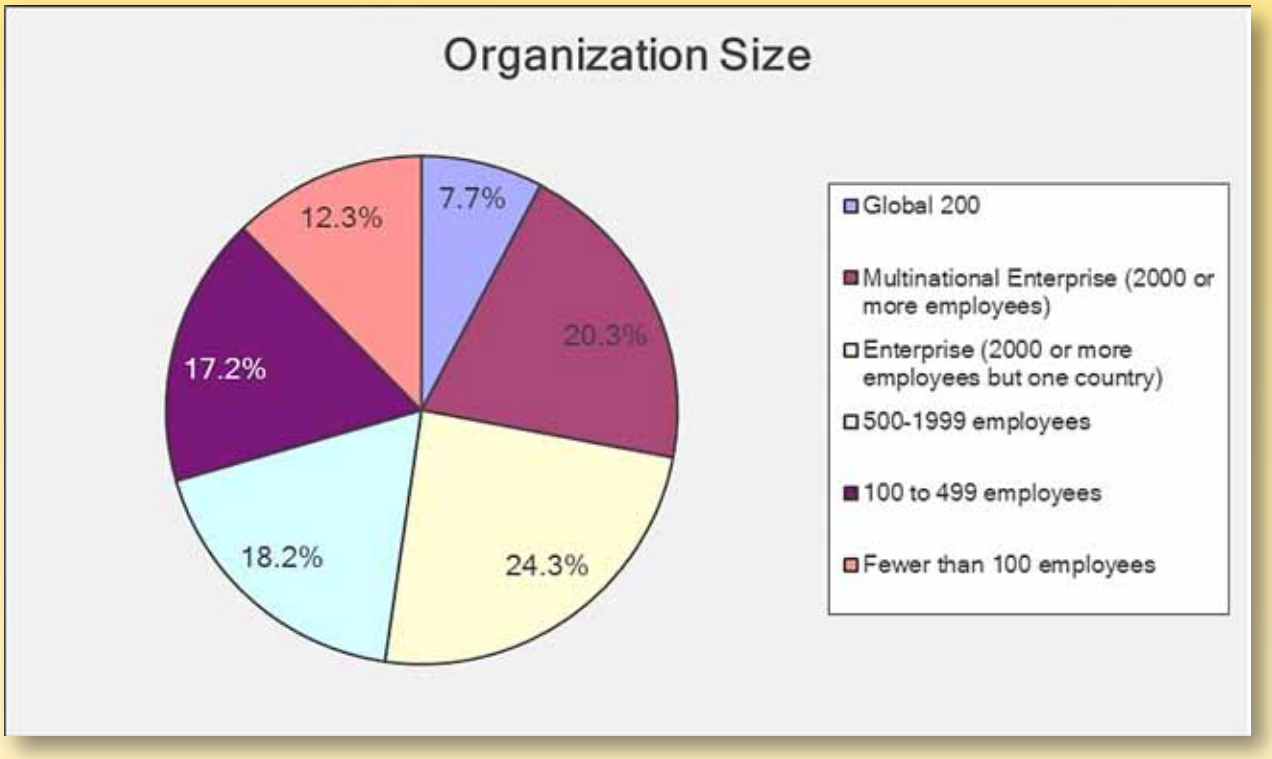

Figure 3. Size of Organizations Based on Responses

The vast majority of respondents held staff positions (rather than being consultants). This year, a higher percentage of respondents held a security-oriented role in their organizations, as opposed to a network-oriented role, which there were more of last year. Of the 747 respondents to answer this question, 73 percent had security titles, whereas 35 percent had networking titles. Some respondents, seven percent, also had compliance officer roles. The total exceeds 100 percent because some respondents' duties overlap among the areas of networking, security and compliance.

Why Companies Collect Log Data

In this year's survey (as in the 2009 and 2010 surveys), detecting incidents, determining what happened (forensics and analysis), and meeting compliance requirements were the top three reasons for collecting logs. Once again this year, the most important reason for collecting log data was to "Detect/track suspicious behavior and prevent incidents," as illustrated in Figure 4. Second place went to "Support forensics analysis and correlation," and third was "Meet/prove compliance with regulatory requirements."

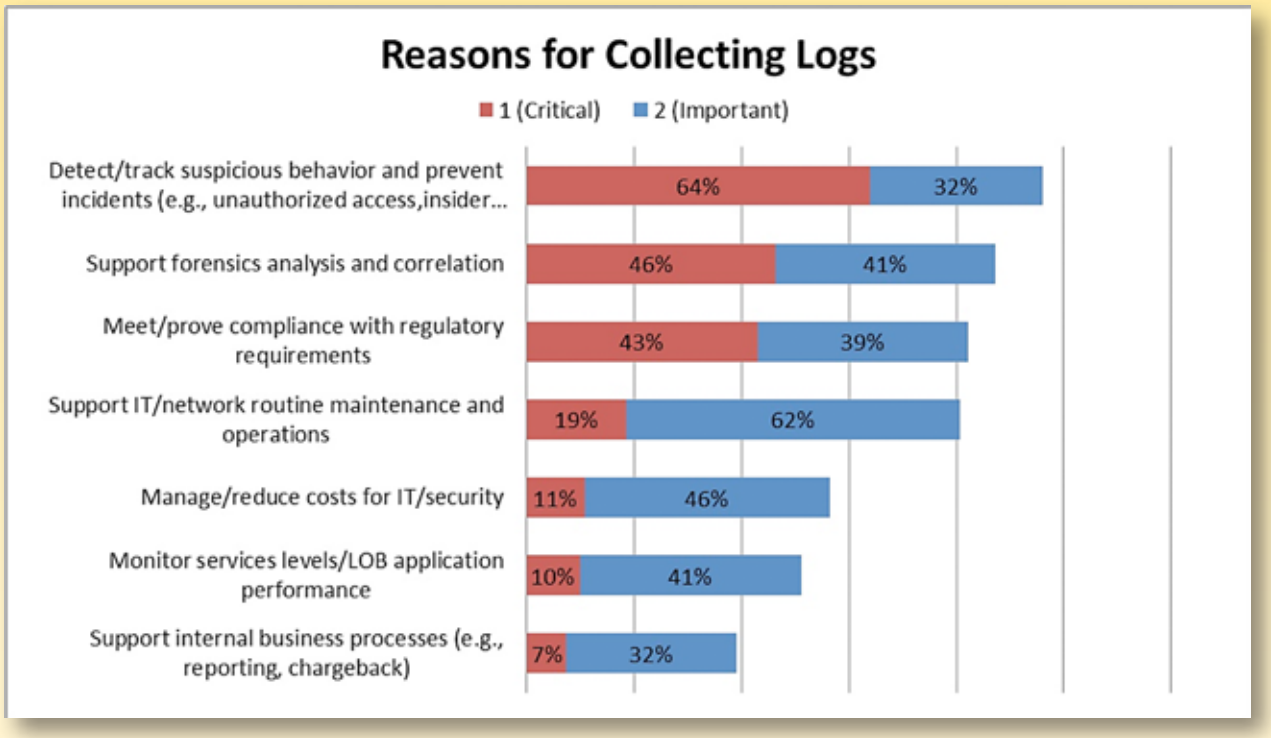

Figure 4. Why Respondents Collect Logs

While maybe not critical, supporting other IT operations ranked high in level of importance, and more than 50 percent of organizations think that logs can be important in reducing costs and supporting other processes besides security and compliance operations. These options were not provided in last year's survey, but survey respondents last year (and this year) indicated an increasing desire to derive more business value from their logs.

Most Useful Features

Once they collect their logs, respondents say the most useful feature of log management systems is "real-time alerts," with 68 percent indicating they are very useful and 25 percent indicating they are somewhat useful. The second and third most useful features were "Intuitive user interface for search" and "Unified interface for all log-related activities." To be precise, there is no such thing as a real-time alert, due to delays in log event analysis and notifications. What's important is that many respondents are getting useful alerts from their log management systems in a timely enough manner.

The fourth most useful feature was "Good performance for all log-related activities, whether individual or simultaneous." In the past, log management system performance received low marks by survey respondents. It is good to see that 55 percent of respondents gave this the highest mark, while 37 percent gave it a mid-range mark. Combined, that's more than a 90 percent approval rating. "Integration with larger SIEM environment" ranked ninth on the list of usefulness. Some comments indicate that respondents are in the process of installing SIEM systems, so there will likely be stronger responses to this question next year. Figure 5 shows the overall ratings for Very and Somewhat useful features based on responses.

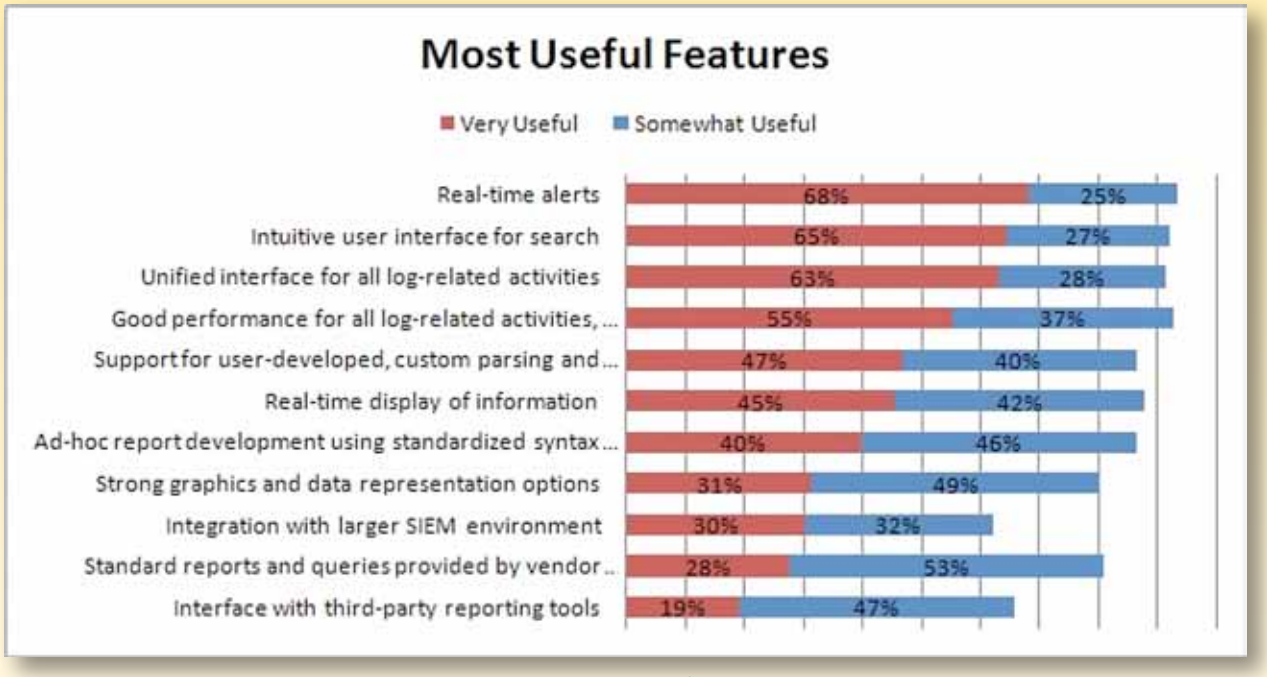

Figure 5. Features Deemed Most Useful by Respondents

Flipping the question around, it's also interesting to note that the least useful features of log management point to other integration problems. The question was, "How useful do you rate the following features in support of your log analysis and reporting activities?" The choices were, Very Useful, Somewhat Useful, and Not Useful. Not Useful was chosen most for "Interface with third-party reporting tools," with 27 percent of respondents choosing this option. Sharing the bottom of the list was with a 21 percent negative vote was "Integration with larger SIEM environment." Figure 6 shows the features deemed least useful by respondents. Overall, these are relatively low negative scores, which suggests that the usefulness of log management systems is improving.

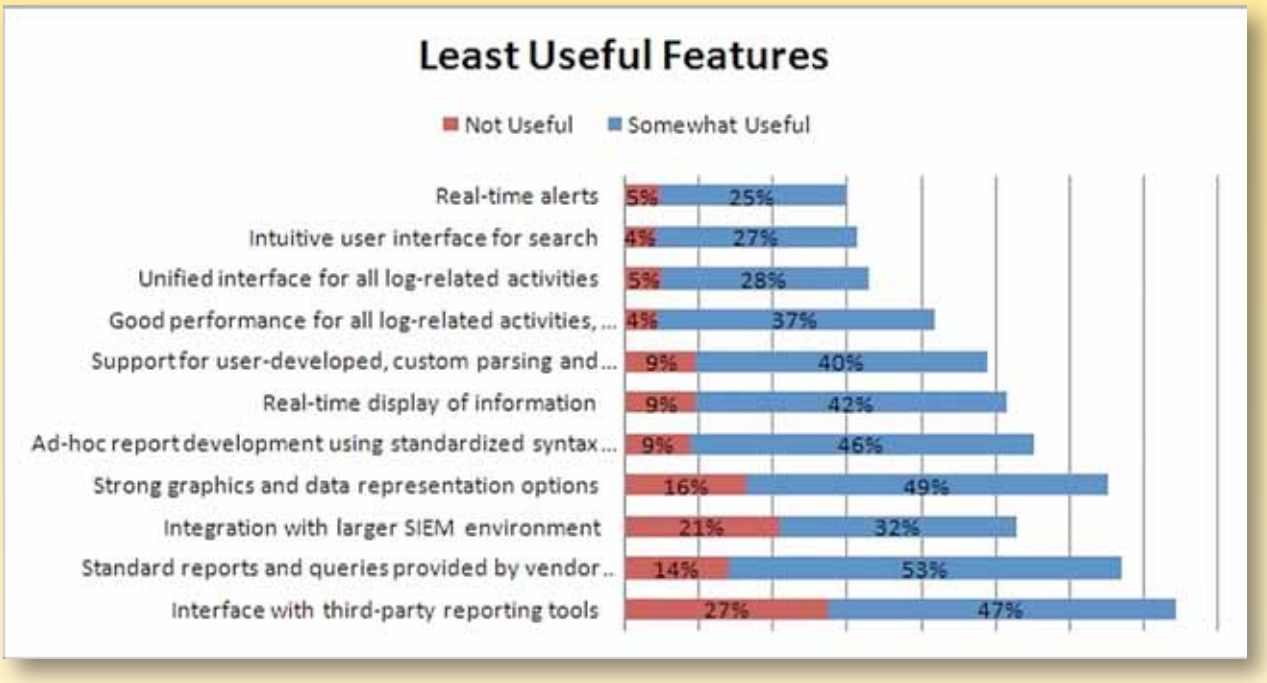

Figure 6. What Respondents Find Least Useful About Their Log Management Systems

Users Want Better Log Data (and More of It!)

The numbers of sources from which organizations are collecting logs continues to expand. This year's survey shows that 59 percent of respondents are collecting log data from their line of business applications, and 14 percent of respondents are collecting log data from their physical plant control systems, such as HVAC. These were not considered a major source for log data in previous years. Other new sources included in this year's survey are log collection from mobile devices (15 percent) and cloud services (14 percent). Point-of-sale (PoS) devices were not on the list but were referenced in comments.

According to this year's survey, most organizations are collecting logs from more than 50 devices, with only 30 percent collecting from fewer than 50 devices. The vast majority of survey respondents indicate they are collecting logs for compliance purposes, leading with PCI DSS. Figure 7 shows what compliance mandates are driving their log management programs.

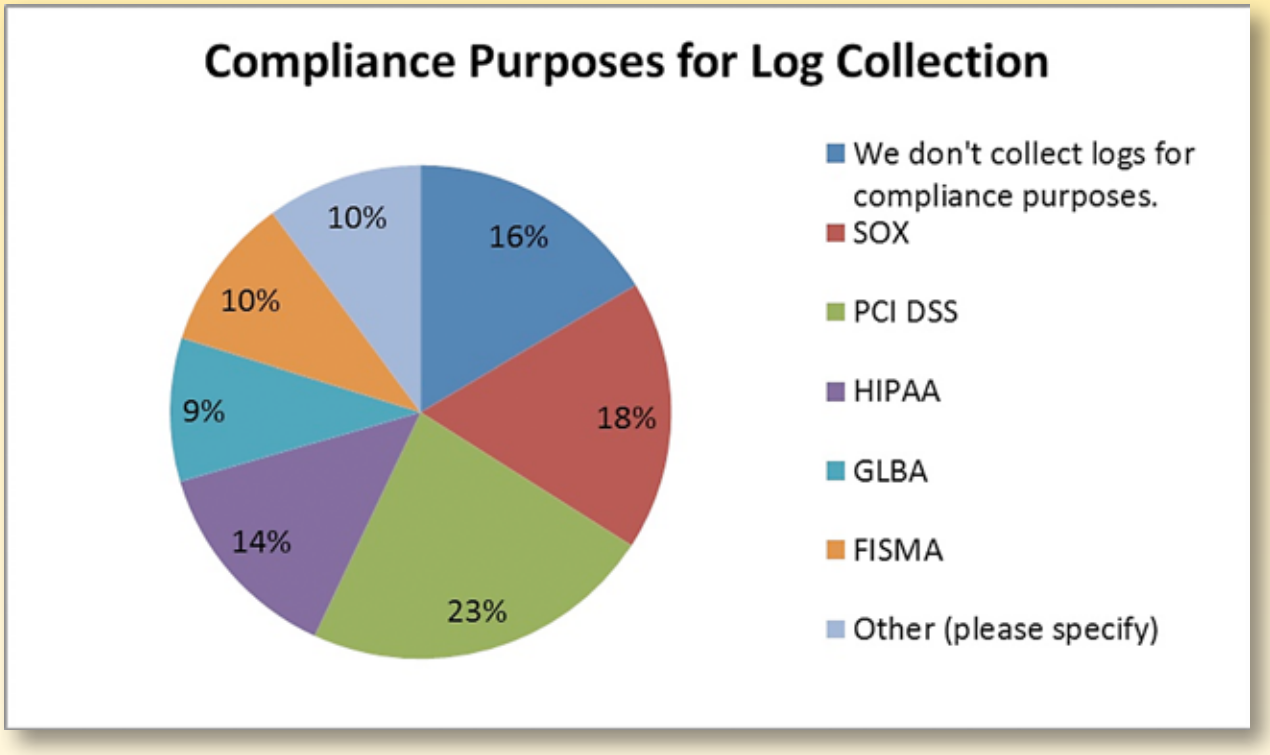

Figure 7. PCI DSS is the Leading Compliance Driver for Log Collection

The types of log information respondents consider to be the most valuable are "Source/destination IP address" and "Time/date stamp." These were nearly tied with "Event information (name, category, type)," followed by "Source/destination TCP/UDP port" and "User information." This level of detailed log data, correlated as needed and in real-time, helps operators find events on the network with minimal manual searching and better accuracy. This question also had an "other" category, in which respondents indicated they wanted even more information from their log management systems, including "detailed network connection logs," "complete URL strings," "full packet capture," and "payload." A log manager might not be the best place for some of that data. Instead, IPS, continuous monitoring or SIEM might collect these data types more effectively. However, the comments highlight the point that many analysts want more information correlated against more threat-monitoring devices to help them make decisions about possible events.

"Vendors need to get better at generating events that are useful because it doesn't matter how good your log management solution is if the events coming into it are garbage," wrote one commenter, Jim Murray, an information security architect in the insurance sector. Vendors of hardware and software that generate logs should differentiate themselves from their competition by standardizing their log data and its syntax and improving the level of log information they make available.

Top Challenges to Effective Log Management

Year over year, trends uncovered in this survey have directly reflected the maturing of the industry. Initially, the top problem reported was simply collecting logs. A few years ago, collecting logs dropped to the least problematic issue, and now respondents express troubles in the areas of normalization, categorization, searching and reporting. See figure 8.

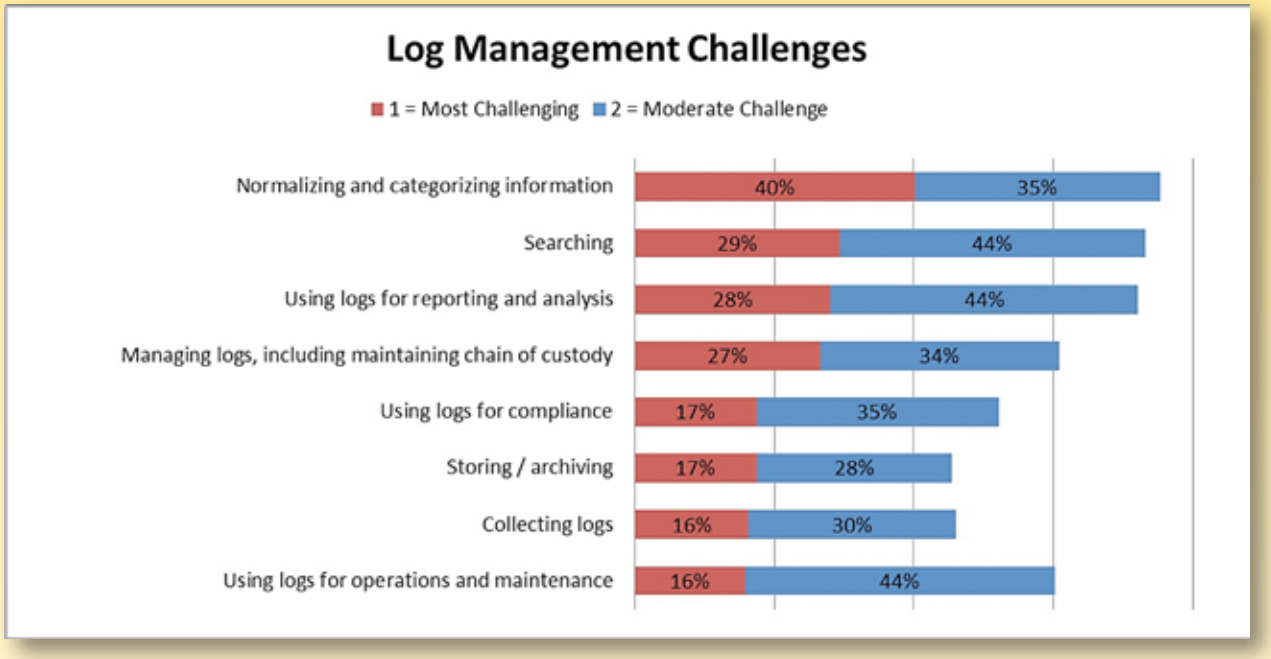

Figure 8. Top Challenges Reported by Log Management Users

"Normalizing and categorizing information" was the top issue this year (42 percent claimed this as their most challenging problem, and 37 percent considered it a problem). The second most noted issue was searching (32 percent considered this their most challenging problem, and 48 percent considered it a problem). "Using logs for reporting and analysis" came in third (18 percent considered this their top challenge, with 50 percent considering this a problem). Nearly as high a percentage (49 percent) considered using logs for operations and maintenance to be a problem, with 18 percent considering it their top challenge. These challenges tie closely to results from a related question about the top hindrances in searching and analyzing logs. In order, these top problems were inability to search across different log management systems, lack of correlation capabilities, interfacing with other IT groups, and locating needed information within the logs collected.

Normalization and Multisource Data

Different systems and devices record the same events in different ways, making analysis of logs difficult. For example, a Cisco ASA firewall, an iptables firewall, and a Check Point firewall basically all perform the same function of blocking some packets and allowing others based on preset criteria. Yet, how they express events in their logs is different for every application. In fact, a Cisco PIX firewall and the newer ASA firewall require some changes to log event analysis when upgrading.

One way to compare similar events is through normalization.[1] Normalization should be able to take log events from all devices under management and present them in a common way for searching and reporting. One problem with normalization is that, unless the log management system saves both the original log data and the normalized log data, the original data is lost to the organization. Original log data can be used for verification and make the difference in determining whether an attack failed or was successful and can point out a false alarm. The problem is greatest when collecting data from systems and hardware (e.g., phones or cloud services) that aren't well supported by the log management vendor.

Most commercial log management systems include storage options for both normalized and original data. These storage systems should be expandable as needs dictate. In the survey, 36 percent of respondents say their organizations store their log data for up to a year, and 33 percent store data for up to five years. Of those respondents who know their log event volume, most see more than 100,000 log events per day—and half of those are seeing more than 1 million events per day.

The bottom line is that each application can log and store different types of data in the formats and for the duration dictated by the organization's business, security and compliance needs. The keys to having good log data management, then, are consistency in format, collection and storage of enough data to answer the "4-Ws" (who, what, when and where), and good documentation to interpret what the log data means.

Getting at the Information

Getting to those "4-Ws" is still somewhat daunting for many organizations represented in this survey. Most log management systems have some sort of a web-based front end that can be used for searching. Responses, however, indicate dissatisfaction with their searching and reporting/analysis capabilities. This coincides with respondents to last year's survey, wherein 64 percent considered searching and reporting to be the first and second most challenging aspects of log management.

When asked new questions about their specific problems with searching and reporting this year, respondents pointed to lack of correlation, inability to search across different log management systems, and integration with other IT systems as their top three hindrances to their searching and reporting capabilities. They also point to problems locating information within the logs.

[1] http://csrc.nist.gov/publications/nistpubs/800-92/SP800-92.pdf

Integration with multiple log management tools is becoming a factor because respondents this year, as well as in recent years past, report using a mix of homegrown and third party tools. Many report using multiple third party log management tools. Responses also indicate multiple homegrown tools in single environments, with a very small number using log management as a service.

Survey responses also point to the need for stronger graphical and data representation, with only 32 percent of respondents ranking these features as "Very" useful in their log management systems. A well-designed graph or chart can convey a lot of information quickly and can even support non-technical managers when necessary. One commenter pointed out that from a business perspective, sometimes including graphics is an expected part of a presentation, even if the graphic's value is limited. Responses indicate that people have worked with their log manager's graphic options and would like to include graphics, but they aren't able to get what they would like out of the presentation capabilities of current log management systems. This is another area of growth for vendors.

The ability to script routine tasks was also brought up by one respondent. Any serious log analyst knows that the ability to set up scripts to run repetitive tasks can be a huge time saver. Scripts often make it possible to track events and statistics, allowing review that would not be available any other way. Many log analysts set up processes to run in the early morning to give them some quick baselines to review when they get into work. Others run scripts periodically to detect suspicious or overtly hostile activity (the single feature rated most useful). In order to collect and consolidate information that doesn't neatly fit into a report, the ability to run low-level scripts is often necessary. Many log management systems have some capability to script and run some reports on a schedule and deliver them over e-mail, via web, pager or smartphone; however, based on responses, they need even more 'scriptability' than they already offer.

Managing Windows Logs

This is the second year the survey included questions specifically addressing Windows log management. The results are essentially the same for both years: Windows, the most heavily used operating system throughout the world, still gets a bad grade for its logging environment. As one respondent stated simply, "Windows makes it difficult to collect logs."

Collection and storage of Windows logs received a 40 percent approval score, with about 10 percent reporting they were "Very Satisfied" and about 30 percent reporting they were "Satisfied." All other categories held dismal satisfaction ratings: Five to seven percent reported being "Very Satisfied" and between 18 and 24 percent were "Satisfied" with their Windows log management capabilities. That leaves approximately 50 to 60 percent of respondents being only "Somewhat Satisfied" or "Dissatisfied" (see Figure 9).

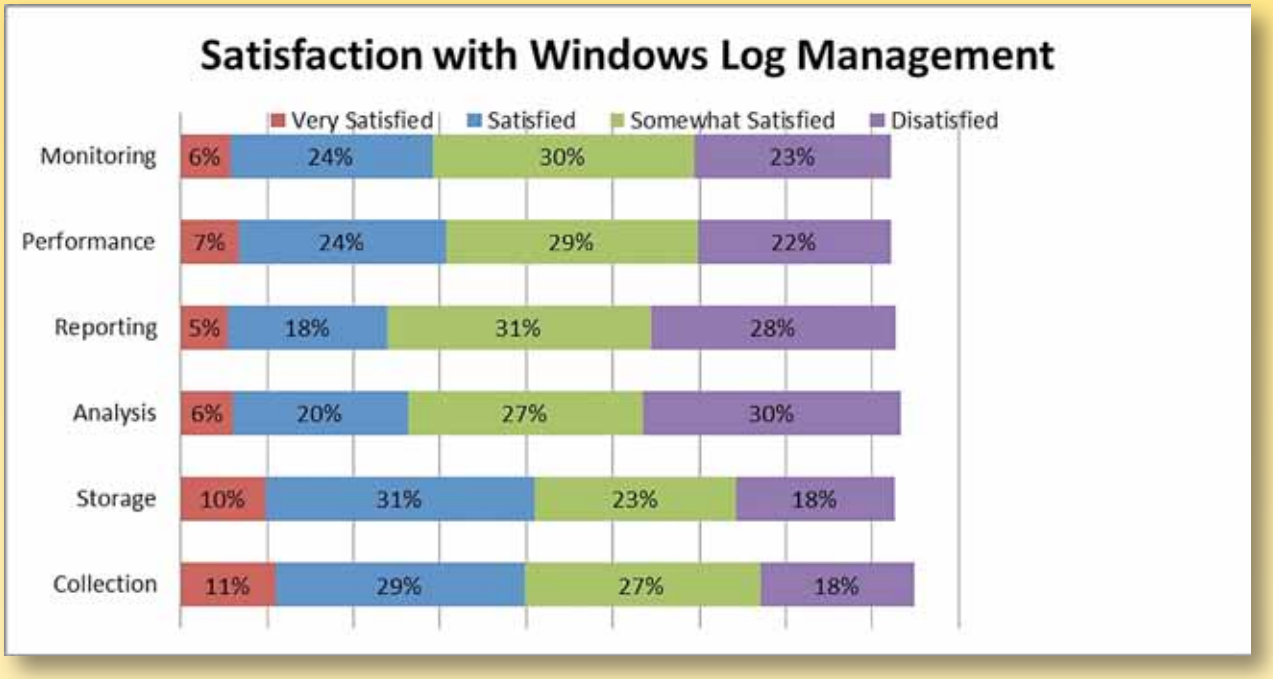

Figure 9. Windows Log Management Still Gets Low Scores from Respondents

Analysis is the top problem that organizations have with Windows log management, closely followed by reporting. There are a number of factors that make Windows log management more difficult than other software (UNIX/Linux) and hardware platforms, such as routers, firewalls and switches. Windows does not natively support syslog in any flavor for log collection. Yet, according to the survey, UDP Syslog is still the most popular log collection method. TCP Syslog is more resilient and can scale better, and 50 percent of respondents also support TCP Syslog. Neither version of Syslog is supported by Windows.

It would be helpful if Microsoft would incorporate some changes in their operating systems to make it easier to collect, normalize, parse and analyze events coming from Windows systems and subsystems. Users often install third-party add-on applications to get this functionality. Those leaving comments listed the Snare agent as the most popular way to send event log data from a Windows server to a syslog server, but there are also other options. Some log management systems pull log data from Windows servers, as well. Today, the burden of analysis rests mostly on the log management software to pull and normalize Windows events into usable information.

Satisfaction with Windows log management has decreased in some categories since last year (monitoring, performance and collection)—with no improvements in reporting and only minor improvements in analysis and storage (see Figure 10). So, vendors have a long way to go to satisfy Windows users.

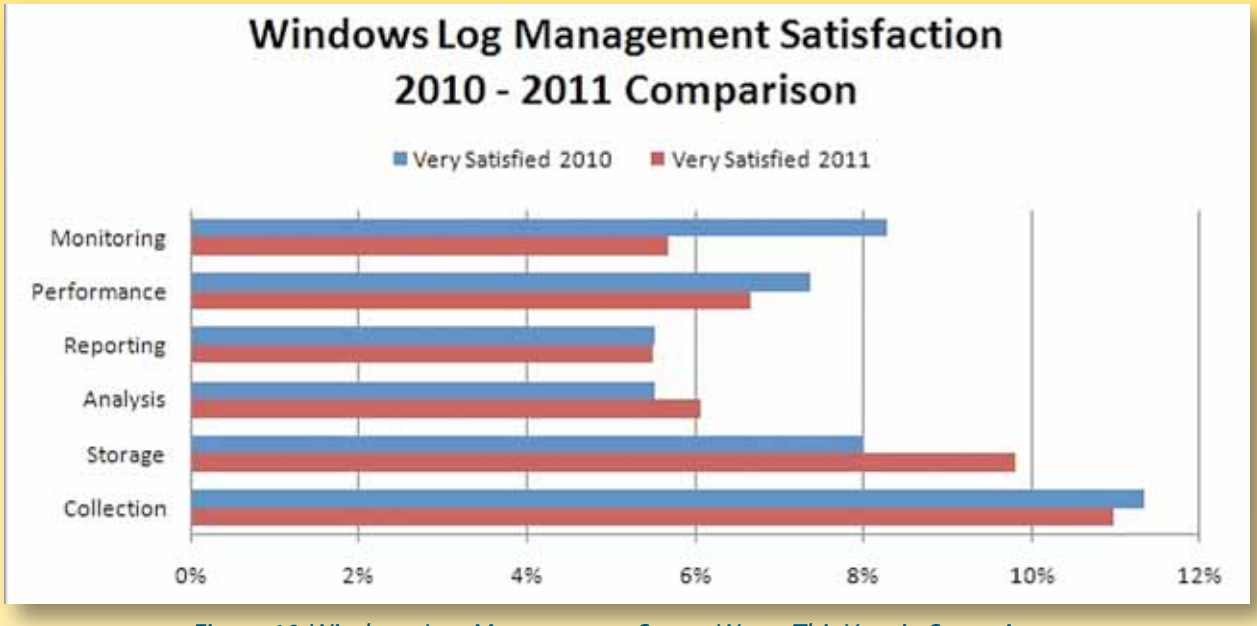

Figure 10. Windows Log Management Scores Worse This Year in Some Areas

Where to Start? A Primer for Windows Log Management

Dr. Anton Chuvakin, lead author of the SANS Log Management course, says, "One of the first things that people should do to start getting value from their Windows event logs is to actually start centrally collecting them from all the Windows systems. Before you can do analytics and alerts, it makes sense to build a working log repository. It will *hugely* help you during incident response."

One popular way to do this is using the Snare[2] agent, although there are other options. It is also possible to pull the information from the event logs using LASSO[3] or one of the other agents that are available.

For a full Windows shop, the logserver could run on a Windows computer. The Kiwi Syslog Server[4] is a popular option. There are also free logservers that run on Linux, and there are a number of commercial logservers. Once the syslog server is running, you can search through the events for events of interest.

Dr. Chuvakin also recommends learning the normal log patterns right after collection. Stored logs are useful (such as for incident response), but to use logs for incident detection, you need to know what is abnormal—and that begins with knowing what is normal!

On the web page for the SANS course on compliance for managers,[5] there is also a link to the course's PDF, which contains a checklist for security incidents. In the lower left corner of that file is a list of a few of the most critical Windows events. These can be a good starting point.

When examining the logs, you'll need a place to look up event IDs to get more information on them. Searching for the specific event ID (e.g., event id 528) on the Microsoft TechNet Support website[6] can be helpful. The site, eventid.net, is also a quick, handy resource for information about specific Windows event IDs. Randy Franklin's website[7] has an extensive list of Windows event IDs.

[2] www.intersectalliance.com/projects/SnareWindows/
[3] http://sourceforge.net/projects/lassolog/
[4] www.kiwisyslog.com/kiwi-syslog-server-features-and-benefits
[5] www.sans.org/security-training/log-management-in-depth-compliance-security-forensics-troubleshooting-1217-mid
[6] http://technet.microsoft.com/en-us/ms772425.aspx
[7] www.ultimatewindowssecurity.com/securitylog/encyclopedia/Default.aspx

Summary

Organizations are increasingly measuring their security effectiveness based on their ability to improve incident remediation, reduce incidents and meet compliance, according to this year's survey. They are also measuring effectiveness by how much they reduce overall security and maintenance costs, as well as improve overall system performance.

Measuring effectiveness and making improvements depends, in large part, upon logs. Log analysts want better log data from more devices, and they are looking for better quality log data to be gleaned from their monitored devices. The top reasons organizations collect logs are to detect, track and analyze security incidents and to meet regulatory compliance requirements. The devices they want log data from are extending beyond the traditional sources (e.g., servers, firewalls and routers) to the physical plant (e.g., HVAC, SCADA) and remotely attached devices, with a small percentage already collecting logs from phones and PoS terminals. IT departments are also looking for log management systems that provide quick, accurate and correlated responses to queries. They also want to be able to turn those queries into reports with visuals and graphics, while being able to easily customize queries to support industry-specific applications and devices in use within their organizations.

While satisfaction is improving overall, respondents are having problems with analysis and reporting. Their biggest problem is managing logs from Windows systems—a pretty big problem because Windows operating systems are so pervasive. In both the 2010 and 2011 surveys, users point to Windows log collection problems and messages that are difficult to analyze. It would be nice to see Microsoft include native syslog capabilities for their operating systems and software. Log management vendors need to continue working to solve the problem, and many are already making headway. IT departments also need to develop internal resources to study log data and learn what events mean. This will take commitment, but the rewards will be increased productivity, compliance and security.

About the Author

Jerry Shenk currently serves as a senior analyst for the SANS Institute and is 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: