SANS ANALYST PROGRAM

Sponsored by LogLogic

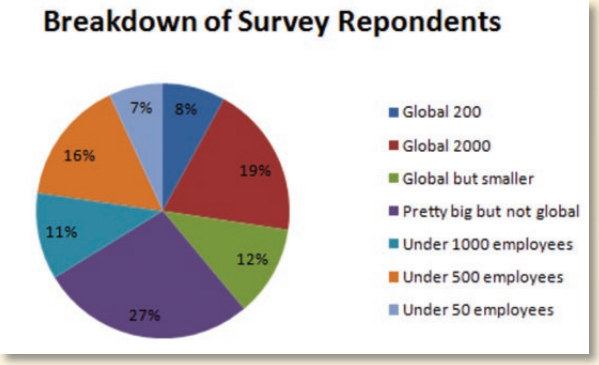
The SANS 2007 Log Management Market Report

A SANS whitepaper

Written by: Jerry Shenk

Executive Summary

The SANS Industry Analyst team conducted its first major survey on the Log Management Industry in the spring of 2005. For the third year, the SANS team has again surveyed IT industry professionals, this year polling them during the Spring 2007 San Diego Log Management Summit. With attendance up, the poll resulted in a much larger sample – 653 respondents vs. just 176 respondents last year. Yet despite the increase in response rate, the results were similar to last year.



Of this year's 653 respondents, the Global 2000 and 200 represent 27 percent, tying with the "pretty big but not global" market size

There has been a slow but steady growth in the number of people who are using their logs to derive value from their log data. In our 2005 survey, 25 percent of respondents were satisfied with their logging situations. In 2006, 28 percent were satisfied. And in this year's survey, 37 percent say they are satisfied with their log information management systems. Clearly, the message about the value of log data is getting out, and people are finding ways to use it.

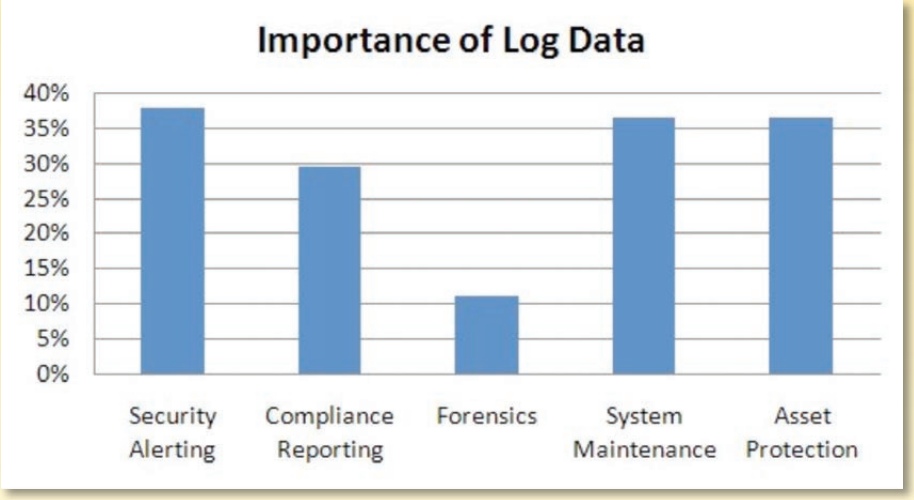
However, with nearly two-thirds still unhappy with their log data management systems, there is plenty of room for improvement, just as there was last year. This is mostly due to lack of correlation and normalization, same as last year. So even though more people are working with log servers and more of them are satisfied with their systems, two out of three are not deriving the information they need from their log management systems. Yet, it's clear from the survey that IT groups want to get more value from their log information. When asked how log data would most benefit their organization, respondents saw 'great benefit' for use of log data in event detection and tracking of suspicious behavior, day-to-day IT operations, process control/compliance, employee use monitoring, forensics, information leak protection and regulatory compliance. When broken down to Global 2000 respondents, regulatory compliance becomes a primary driver.

How they're using the data, however, is a different story. Most are not using their data for forensics and compliance, despite their importance. And two-thirds (63 percent) of respondents are not delivering reports including log data to their executives and managers.

This paper analyzes the survey data to unlock how log data is being used successfully, the key problems holding enterprises back from log management, what's still needed from the vendor community, and how vendors are working to resolve these issues. A Technical Addendum on how to start interpreting your log data has also been included.

Importance of Log Data

In this year's survey, respondents were asked to rank the three most important types of log-related activities to their organization in order of first, second, and third choices. Leading their first choices was Information Asset Protection (46 percent), followed by system maintenance (35 percent), with security monitoring and compliance tying for 31 percent. Dominating their second choice was compliance reporting (36 percent). Interestingly, when you scroll over to the third choices, forensics, which is a dead last for actual log data usage in our survey, is the first choice in terms of data importance. Yet despite their importance in second and third choice categories, neither compliance nor forensics is among the main uses for log data today. This is indicative of several things, which we will cover later in this paper.

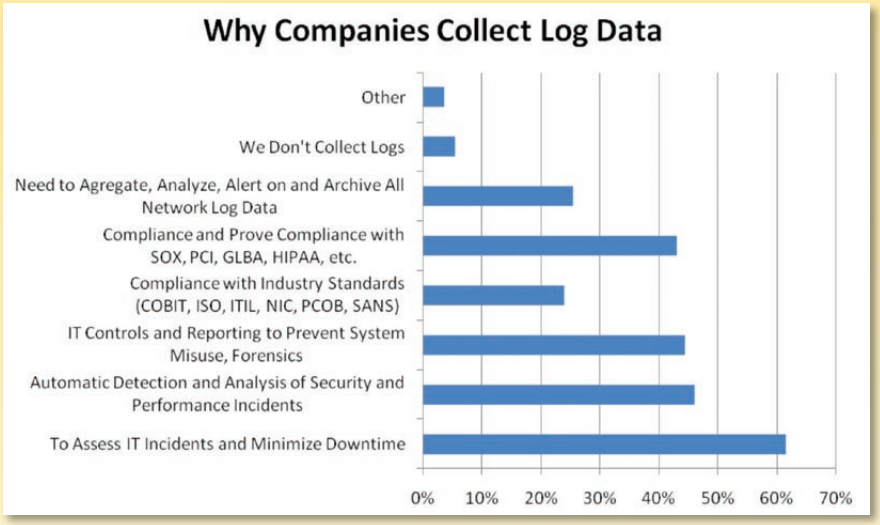


Most important types of log data, averaged across three choices.

When you add up the number of respondents across their three choices (see above graphic) you get a slightly different picture. Looking at it this way, overall the data shows that respondents deem security alerting (38 percent), system maintenance and information asset protection (tied at 37 percent) and compliance reporting (30 percent) most important. And, like last year, forensics clearly lags with only seven percent of respondents considering forensics to be a driving factor in maintaining log information.

How Is Log Data Used?

How respondents use logs is often different from why they're collecting log data, according to the survey. The number one reason they're collecting log information is to have it on hand to be able to accurately assess IT incidents and minmize network downtime. Some 62 percent of respondents say this is their reason for log collection. The second reason people gave for collecting logs was automatic detection and analysis of security and performance incidents, indicating they are tying their logs into a Security Information Management (SIM or SIEM) and / or intrusion prevention systems.



System maintenance was given as the number one reason for collecting log data, which feeds into the next reason: IT controls and compliance, each of which produced similar statistics among respondents at 46 and 44 percent respectively.

System Maintenance

No system administrator should be at all surprised that System Maintenance was ranked as a top use of log data. Based on the entire sampling, 62 percent say they collect log data to minimize downtime and assess IT incidents. In fact, the lack of log data is a serious obstacle to overcome when attempting to resolve system problems. End users often regard alerts and warnings as nuisances so they close those messages without recording the information. If the data is not logged someplace, it is necessary to re-create the error so that the support personnel can see exactly what happened. The availability of a full complement of log data from the application to the workstation, server and infrastructure can fill in the details that the end-user may not even be aware of.

Larger Companies Differ

When we did the breakdown, the larger companies revealed the same basic trends overall, although there was a noticeable increase in the use of log data for compliance-related reasons such as SOX and PCI DSS mandates. There was also a slight decrease in using the data for minimizing downtime/system maintenance.

Breakdown of Global 2000 and 200 reasons for collecting log data

Archiving

When you look at just the global 2000 responses to this question, the two top reasons for collecting data are archiving and compliance reporting, which are obviously related. Yet, based on their storage retention uses, organizations are not maintaining these logs indefinitely for compliance purposes. Most (14 percent) are unsure of how long they maintain their logs or they rely on the O/S default for that system. Just over 11 percent store their log information for 30-90 days, and a mere nine percent store their log data for six months or more. This is due to many factors, not the least of which is the sheer volume of data these systems produce and their lack of common format.

Compliance Reporting

Compliance reporting is also a growing concern among respondents. In fact, our research put it exactly on par with security alerting and reporting. Over the past few years, regulatory bodies have considerably increased the requirements for logging of security-related data. Much of the data today required by regulations goes well beyond logs from their network and security devices. It also includes managing log data from applications where sensitive data might be stored and accessed by end users. This includes operating systems, databases, home grown and commercial applications, and mainframes. Tracking access to restricted data must become part of normal operation, as should the ability to tell when there is misuse of access to data.

Survey respondents are collecting this data to varying degrees. Most (79 percent) are collecting firewall log data. After that, other forms of data collection drop off precipitously. Collection of antivirus, routers and IDS/IPS is done by 57-58 percent of respondents. And at the application level, 57 percent are using their O/S logs, 55 percent are using their database logs, 49 percent use logs in their enterprise applications, 31 percent use logs on home grown applications, and mainframe application logs are used among 23 percent of respondents.

Using Log Data Forensically

In the survey, the forensic use of log data was ranked as one of the top three important imperatives for IT organizations. Yet it also ranked substantially lower than any of the other choices as a "chief" reason to collect log data. Only seven percent of respondents chose forensic use as their top reason for storing log data. One reason that could factor in for low forensic use of log data is that incidents rarely occur that require the use of forensics, which is a legally-rigid process of digital evidence gathering. As disclosure laws are strengthened on organizations housing or processing personal consumer data, we expect the demand for log data in forensics to rise over the next few years.

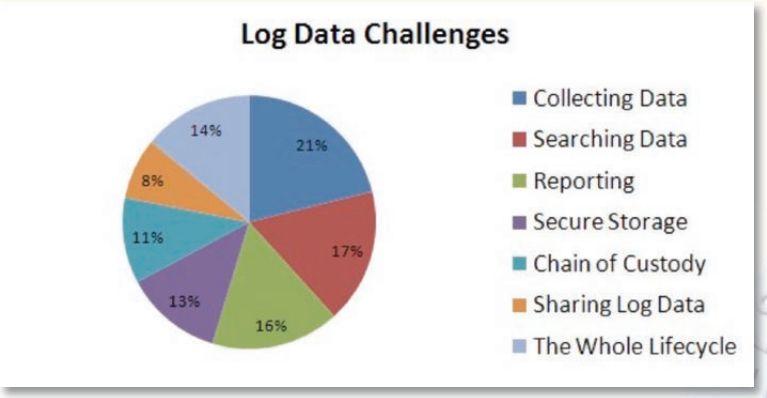
Why Don't People Use Log Data?

Based on the 2007 survey results, IT managers realize the usefulness of the log data floating around their enterprises. They also know that these logs hold valuable information that, when unlocked, can serve a variety of risk management, regulatory and assessment functions, as noted earlier in this report.

To this end, 40 percent of survey respondents review log data once a day or even more frequently. And there was a large gain in the percentage of companies reporting they had log servers. In this year's survey, 57 percent report having log servers, vs. 35 percent last year. However, 43 percent of them are still not using log servers to better realize the potential of their log data. And even those that do have log management capability are not satisfied with their systems. As the survey shows, a clear majority (63 percent) of people are not satisfied with their current log file analysis processes. This shows an improvement of nine percent over last year, in which 72 percent of respondents were unsatisfied with their log file analysis.

The main reason for the continued dissatisfaction is that the enterprise users know that they have a difficult task in trying to build or buy a comprehensive log data management program for collecting, searching, condensing and reporting on the vast, rich sources of log data across their enterprises. This is due to a variety of reasons, not the least of which is the sheer volume of log information being produced by their routers, switches, applications, operating systems, mainframes (and, in some cases, physical security data where integration occurs). That's because, like last year and the year before, the primary problem continues to be the lack of interoperability between log information sources.

In the area of challenges related to log data, respondents were asked to rate seven choices based on criticality in a multiple choice question. Top problems selected by respondents included the collection of log data (27 percent), searching log data (22 percent) and reporting on log data (21 percent) all statistically tying for second. Other challenging issues included secure storage (16 percent), chain of custody (14 percent) and sharing log data (10 percent). And 18 percent of respondents indicated that the whole lifecycle was a problem.



Multiple choice question on log management problems, averaged out.

Collecting Log Data

The top complaint from the 2007 SANS survey respondents had to do with log data collection. Over one-fourth of the respondents stated that this was their most critical problem. Some systems have built-in ways to export log data, yet virtually every system has its own way of formatting data that is logged. Some of them are easily processed by computers because of consistent formats, but most log formats are inconsistent with one another, and therefore can be quite difficult to process automatically. Even worse, many applications don't have any logging capabilities at all, based on hands-on analysis with healthcare, billing and phone systems.

Large applications like e-mail servers and SQL databases have options for logging. Most server operating systems and infrastructure components also have substantial logging capabilities. These systems require configuration for logging because they ship with only minimal logging capabilities enabled by default.

Syslog servers, too, have evolved as the de facto standard with 72 percent of respondents indicating that their log servers are primarily using syslog. Syslog is a defined standard, documented in RFC 3164[1]. Syslog uses UPD port 514 for transmitting data. The syslog protocol is not perfect – there are some limitations such as plain text log data on the network, and unauthenticated log devices. One possible alternative is Syslog-ng, a plug-in replacement for syslog but with better network connection support (TCP vs. UDP) and a richer set of configuration options.

SNMP traps are another method for delivering log data to the log server. SNMP traps are transported using UDP port 162. The related RFCs are 1157[2] and 1215[3]. Some network devices use SNMP traps instead or in addition to syslog.

Using Log Data – Searching and Reporting

In last year's reports, we stated that "…many IT people ignore their logs because it's too much work to manage them." Andrew Davis of University of California, San Diego said, "Rapid evolution of our enterprise IT infrastructure has resulted in exponential growth of data." This is particularly true when you consider how difficult it is to convert data to a single readable format.

This is because vendors have created different ways of logging events. Windows and Novell servers typically store their log data on the server at the point where the event occurred with no easy way to send the data to a log server. While there are ways to work around this issue, it is an area that needs attention at setup time.

[1] http://www.ietf.org/rfc/rfc3164.txt
[2] http://www.ietf.org/rfc/rfc1157.txt
[3] http://www.ietf.org/rfc/rfc1215.txt

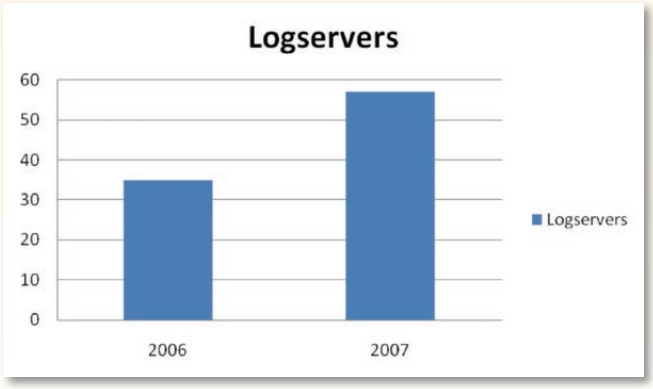
The lack of standard log formats is also inherent in single vendor products and log formats that sometimes change with each upgrade. If a company was using PIX firewalls two years ago and they've upgraded to ASA firewalls, the log data has changed only slightly but enough to warrant changing automated processing to match the changes in the data itself.

If you have a large environment that covers different time zones, this also causes complications and errors when converting the data. Even on a single computer, some applications may store data across a variety of different time zones. For example, most Microsoft Windows system logs are stored in the time zone of the local computer, but Microsoft IIS web server logs are stored in UTC format. UTC stands for Coordinated Universal Time and is also known as Greenwich Mean Time or Zulu Time.

There is a real market opportunity here for vendors to step in and create a CVE-like body that standardizes these formats and creates products with adequate log analysis coverage in ways that simplify log management for today's diverse enterprise.

What's Still Needed?

Consumers of log data information are still not getting their log information digested in useful enough ways, as indicated by the fact that 63 percent of respondents are not passing data on to other groups in their organization. Another indicator is that 63 percent stated that they were not satisfied with the data they were getting, as mentioned earlier.



More people are using log servers in 2007

Log Data Normalization

One solution to the lack of log data consistency is normalization. Normalization is achieved by having a log server collect the data from the various sources and then convert them to a common format, either at the time of collection or at the time the information is delivered up to the user interface. Normalization helps improve event correlation but often causes the loss of some degree of detail. If normalization is required, it's best to continue storing log data in raw format so that the details are available if needed at a later date or for compliance or litigation reasons, both of which require that this data be untouched and copied as a mirror image in examination cases.

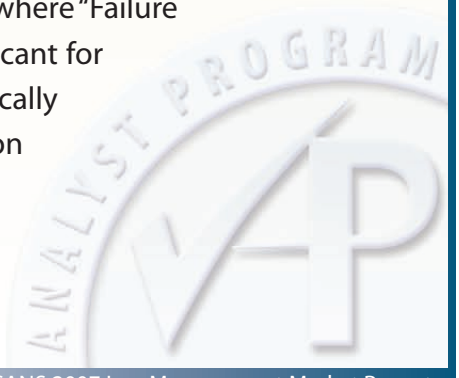
Infrastructure Logging

One area of logging that is often ignored is the logging of the network infrastructure. This would include switches, routers and firewalls. In the survey, 50 percent of respondents indicate that they are logging switch data. Switch data can be used to determine when machines were turned on and off, they can provide indications of power failures, and can often point out network configuration errors. Switch data isn't typically the first place to go when there is a problem, but it can be used to provide additional information that can be helpful in determining what machines were turned on and when. If events point to a specific machine being the source of a problem, tying in logs from the switch can provide additional evidence that it was or was not the machine in question.

System Planning and Early Warning

Log data can also be used for system planning by identifying trends and reporting system errors. For example, by monitoring the Windows event logs for "disk is at or near capacity," it may be possible to detect a shortage of drive space prior to actually running out of disk space. There are numerous other methods of monitoring for low disk space but this is one way that doesn't require additional software installation.

Most operating systems start generating errors long before operational problems are noticed. Some other key indicators of problems are the "red X errors" in the windows event viewer. The "red X" is shown in the windows event viewer but the word "Error" followed by a server name is a very good indication. This is also the same location in the data where "Failure Audit" would indicate an Account Logon failure. The case is significant for both this error type and the server name. And, this location is typically the 10th field in the log data, but it should be at a consistent location for each log server.



Syslog Compatibility

Every operating system vendor, application vendor and software vendor seems to do things just a little bit differently when it comes to generating log data. Clearly, with 72 percent of respondents having syslog servers, customers need their log management products to have the ability of syslog for message transport. Yet many vendors still don't support this.

Log management would benefit if vendors made their products consistent in their logging data formats. Application software vendors also need to get on board by generating basic, useable log data, starting with login and logout data, including failures and successes. Failed attempts to access data could indicate the malicious guessing of passwords from unauthorized users. Successful logins can indicate that not only did the guessing work for access, but also who's logged on, where, and when. Other logged information, such as changes to user information, access lists and permission tables are also critical pieces of information telling of a likely intrusion.

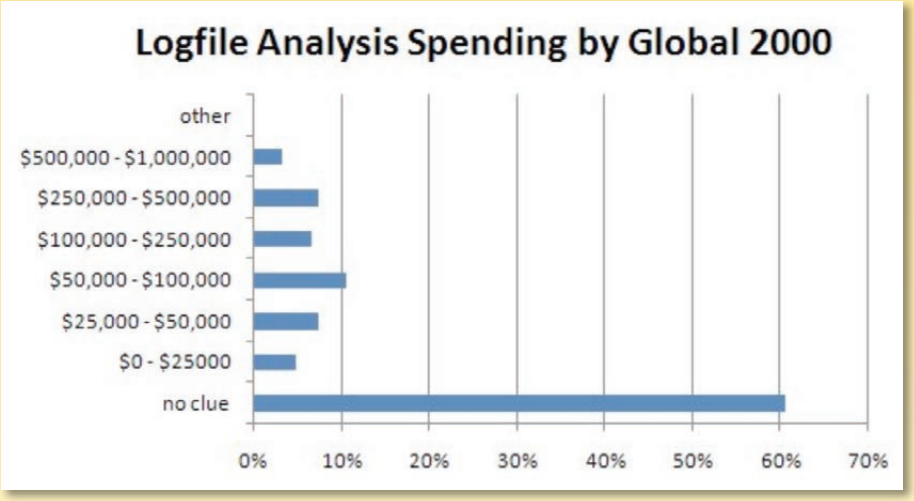
One other helpful piece of data would be to log communication failures. On a recent, multi-site IVR (Interactive Voice Response) system, the only way to know that a communication failure occurred was that the system would not ask the caller to leave a message. A company doesn't have any way to know if their voice mail system is down unless someone tells them. A little bit of log data would make life much simpler.

Moves in this direction are happening. There is currently a group working on bringing some standards to logging. Security expert Dr. Anton Chuvakin has recently announced CEE (Common Event Expression) on his blog[4]. This working group has been established to work toward "A standard log language for event interoperability in electronic systems." This is a new group so it will be some time before we see the fruits of their labor. And the government is looking at standardizing efforts – NIST recently published Special Publication 800-92: Guide to Computer Security Log Management[5]. One of the problems with log management cited in the report is the inconsistency of log data from various sources.

[4] http://chuvakin.blogspot.com/2007/04/finally-common-event-expression-cee-is.html
[5] http://csrc.nist.gov/publications/nistpubs/800-92/SP800-92.pdf

The Log Management Market

Global 2000 companies spent an average of $187,000 annually on log management, according to the SANS Log Management Survey. This equates to a $374 million annual market. This averages out to about two hundredths of one percent (.02%) of profits based on the Forbes 2000 list[6]. These statistics are similar to those on the 2005 and 2006 surveys.



What the G2000 (including the G200) are spending on logfile analysis.

There are quite a few companies that deal with security logs and system monitoring but few that focus on log management itself. There are managed services companies that monitor system and security logs as a service, and there are a number of commercial and open source programs that monitor performance metrics, but most log management is being done with homegrown solutions. This is down from the surveys in prior years, indicating that more companies are using vendor-provided solutions. Among the larger companies (the Global 2000), 55 percent are using an appliance-based log management solution. Many times, different solutions are linked together to provide a solution.

[6] http://www.forbes.com/lists/2007/18/biz_07forbes2000_The-Global-2000_Rank.html

Log Management – Where Do We Go From Here?

Today, over half of the Global 2000 companies (55 percent) report they are not happy with their log management solutions. As evidenced by comparing the 2007 survey results with those of prior years, the market has grown; but the surveys also indicate that there is still room for improvement. More people are working with log servers and more of them are satisfied with where they are but nearly two out of three indicate they are not satisfied with their log management systems.

Two key problems with log management are collecting data, and processing and reporting that data. These issues naturally stem from the sheer volume of log data, inconsistencies in log formatting and the lack of available logging features in many applications.

Vendors of log management systems are working to resolve many of the issues raised in this year's survey but in the meantime, IT staff members need to take it upon themselves to learn about their systems and the value of their log data (see our Addendum: Understanding Syslog Data). As vendors, working groups and users solve the problems associated with log data management, we will see a more mature market in which everyone benefits from their important log data.

Addendum 1: Understanding SysLog Data

Sometimes, the best way to learn a system is to set someone down in front of the logs and let them read them. That may sound like a daunting task at first, but with a little practice, they really aren't that bad… Ok, they are pretty dry reading, but if you take the time to study the logs, you can soon start pulling valuable data out. If you can learn to use the tools included with your log server, you can quickly learn to extract the important data and ignore the redundant, repetitive and uninteresting. "Uninteresting," means unrelated to a current task. Note that "uninteresting data" for one analysis may be critical in another analysis.

Finding Interesting Data

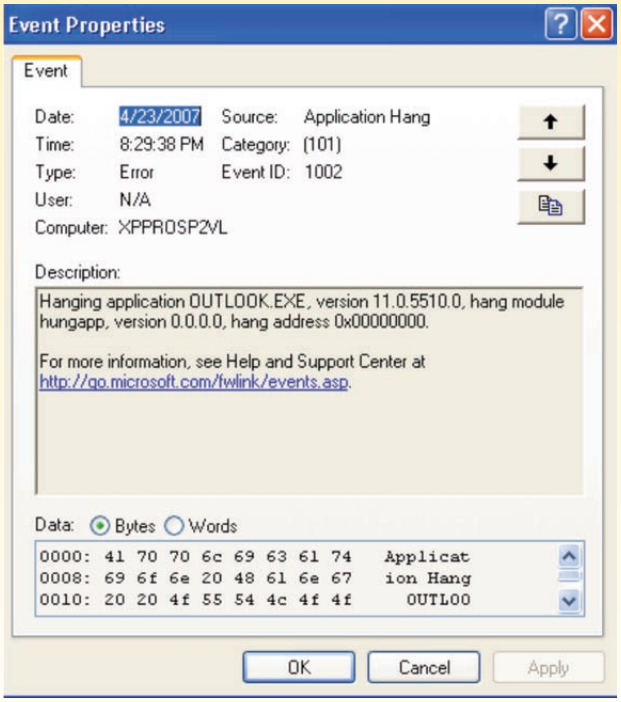
Since 72 percent of people taking this year's survey are primarily running syslog based log servers, let's look at a two specific examples of ways to work through the data on a Linux or UNIX based syslog server. If you can work through these examples, you can then apply the same technique to other log data.

While these examples are done using a tool called Grep on a Linux or UNIX based system, the same logic applies to any log server or log management appliance. You can search for data you want, and you can exclude data you don't want.

Understanding Windows Log Data

Randy Franklin Smith has a website dedicated to Windows Security and a large part of that is making sense of Windows Event Log data. One page in particular, "Randy Franklin Smith's Security Log Encyclopedia,"[7] has a quick list of Event IDs, which Windows version they relate to and a brief description.

The Microsoft TechNet website can be a valuable resource if you can find what you're looking for. Their Security Monitoring and Attack Detection page[8] ties right in with the "Top 5 Essential Log Reports." When viewing events in the Event Viewer, most of them have a link to the "Help and Support Center" that will display a description of the error and possible actions to resolve it. This screen capture image shows an error generated when testing the recent Windows Animated Cursor Buffer Overflow exploit[9].



A Windows log event capturing an ANI cursor exploit.

[7] http://www.ultimatewindowssecurity.com/encyclopedia.html
[8] http://www.microsoft.com/technet/security/midsizebusiness/topics/serversecurity/attackdetection.mspx
[9] https://www.sans.org/webcasts/show.php?webcastid=90861

Understanding Cisco Log Data

By default Cisco routers, switches and firewalls have a minimal amount of logging enabled. It is the norm across the industry that the default level of logging is minimal if it's enabled at all. There are differences in how logging is set up between routers, firewall and switches. To add to the confusion, there are even a few varieties of switches. In most cases, using the command **logging on** will enable logging. Then the command **logging buffered debugging** will cause quite a bit of log data to be stored to an internal buffer. That internal buffer will be cleared if the device is rebooted. To send basic data to a log server with an IP address of 10.1.1.6, enter the command **logging 10.1.1.6**. The command **logging trap debugging** will cause the device to send a lot of data to the log server with a syslog listener. The syslog listener should be listening on UDP port 514. In some cases, where there is an extremely busy network, debug level logging may be too much; but many engineers would rather have more log data than they need rather than risk not having enough.

Cisco log data is quite different for different devices. In the following example, we can see a PIX firewall starting up, a successful SSH login for the user PIX and a configuration change to enable logging (obviously, logging was already enabled or this information would not have been logged). In this example, the first timestamp is that of a Linux-based syslog server. The second timestamp is the timestamp from the PIX firewall.

```
May  3 18:59:12 10.1.1.254 May 03 2007 16:05:01: %PIX-6-199002: PIX startup
completed.  Beginning operation.

May  3 18:59:54 10.1.1.254 May 03 2007 16:05:44: %PIX-6-605005: Login
permitted from 10.1.1.125/4434 to inside:10.1.1.254/ssh for user "pix"

May  3 19:00:40 10.1.1.254 May 03 2007 16:06:29: %PIX-5-111008: User
'enable_15' executed the 'logging buffered debugging' command.
```

In another example of Cisco log data, we have data from a router. In this case, we can see that a configuration change has been made from the console and we can see the logging command that's listed above.

```
May  3 19:01:49 10.1.1.211 7529: *Apr 14 08:20:08.819: %SYS-5-CONFIG_I:
Configured from console by vty0 (10.1.1.125)
```

One valuable resource for interpreting output from Cisco devices is the Cisco output interpreter[10], available from the Support Tools and Resources page. This page will require a valid Cisco account. It is possible to paste the output from a Cisco device into the web page for interpretation.

If you're just getting started with log analysis, you will do well to target your initial work in the areas where you will get the most value. This case study is an example of three critical security areas: powerful accounts, powerful systems and vulnerable systems. A "powerful system" (DC Role) had a "Vulnerable System" (DNS) running as a "powerful account" (SYSTEM) that resulted in a full system compromise.

[10] https://www.cisco.com/cgi-bin/Support/OutputInterpreter/home.pl

Powerful Accounts

The SYSTEM account and Administrator accounts are the most powerful accounts on a Windows system. Comparable accounts on a UNIX or Linux computer would be the root account. Novell systems would normally have an Admin account. In any given environment, those accounts could all change so it's important to tune your log monitoring efforts to suit your system. There may also be accounts with access equal to these named accounts.

Other powerful accounts are those of network administrative personnel. Even if these administrators have limited access, they can still do a lot of damage. The damage could be done by an outsider compromising those accounts, an insider who learns a password – or even the administrator himself. By logging what network administrators do, it is possible to reconstruct events leading up to a compromise or determine what mistakes were made in the event of an accidental incident.

Accounts belonging to managers or human resource personnel are another group of accounts that should be monitored. Monitoring these accounts, which helps keep people accountable for what they do, is also recommended by many regulatory bodies including HIPAA, SOX and PCI.

Powerful Systems

The most valuable systems in a Windows environment are the servers with the Domain Controller role. Other valuable systems would be systems with critical data. But if a Domain Controller gets compromised, then every other system including those with critical data is vulnerable.

In any environment, systems with critical or highly confidential data should be included under this definition of "powerful systems."

Vulnerable Systems

Vulnerable systems are harder to identify. A fully patched server may be vulnerable to unknown issues. At the end of March 2007, an Active Directory server with the Domain Controller role and a DNS server would not have been considered particularly vulnerable. It was very valuable and these servers should have limited access but the OS was thought to be fairly secure. Even when the news broke on April 12th about the RPC/DNS vulnerability[11], it wasn't immediately understood just how critical it was. Even at the end of April, many administrators didn't realize that it could lead to a full system compromise.

One common classification for vulnerable systems is when access is available without accountability… typically meaning from the Internet. Monitoring the logs of Internet- accessible systems is critical because of the likelihood that there are vulnerabilities in the operating system and applications that may not yet be publicly known.

If an organization has wireless Access Points, they would also fit this description of "vulnerable systems" and should get special attention from a monitoring system. If access points are configured to send events to a log server, the recent attack in which de-authentication packets are sent repeatedly in an attempt to crack the WEP key[12] generates a lot of log traffic. Even if only statistical analysis is being done on that log data, the high volume of these log entries should raise an alert after the attack. Now that the attack is known, it can be specifically monitored.

[11] http://www.microsoft.com/technet/security/advisory/935964.mspx
[12] http://radajo.blogspot.com/2007/04/what-else-do-you-need-not-to-use-wep.html

About the Author

Jerry Shenk

Jerry Shenk currently serves as Senior Analyst for the SANS Institute and is the Senior Security Analyst for D&E Communications in Ephrata, PA. Since 1984, he has consulted with companies, 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 5 GIAC GOLD certifications: GCIA, GCIH, GCFW, GSNA and GCFA: all completed with honors.