SANS ANALYST PROGRAM

Sponsored by ArcSight

ArcSight Logger Review

A SANS Whitepaper – January 2009

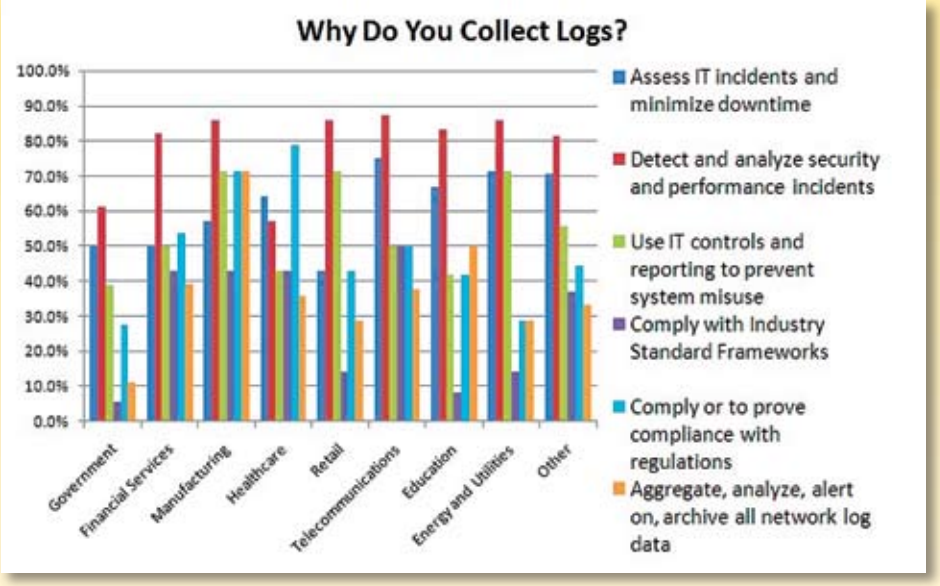
Written by: Jerry Shenk

Getting Started: A Requirements Checklist

Review: ArcSight Logger 7100 v.3.0

Executive Summary

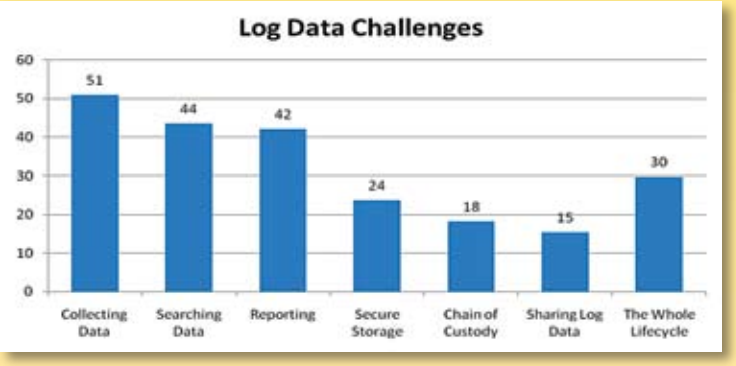
Over the past four years, the SANS Annual Log Management Survey has shown increased interest in log management as organizations begin to realize the value that their logs can provide to auditors, security teams, network administrators, and even operational business units. Compliance with PCI, SOX, HIPAA, and other regulations is the primary driver for collecting logs, according to 78 percent of respondents in one part of the survey. But another question showed that logs are also being viewed as a means to increase visibility into networks, improve overall security effectiveness, and minimize downtime.



Source: SANS 2008 Log Management Survey[1]

Organizations collecting these logs now know they have access to all this rich, business- and security-enabling data. But how do they manage those logs? How do they parse, normalize, organize, search, get reports from, and store all of the data in a friendly, useful, and scalable manner? A significant number of survey respondents indicated frustrations on all these levels of log data usage. This leads to the conclusion that, while they may be collecting their data for these purposes, they're not getting the valuable usage they require.

[1] www.sans.org/reading_room/analysts_program/LogMgt_June08.pdf



Source: SANS 2008 Log Management Survey[2]

It's clear that log management systems must do more for their enterprises in all these areas. So, with these drivers in mind, we tested the ArcSight Logger 7100 v.3.0, which, overall, handled an enormous amount of logs from a variety of sources with ease and analyzed them simultaneously.

This version of Logger collects data from 275 log data sources and has 4.5 TB of physical storage for raw and normalized data (which translates to 35 TB of effective capacity, according to product information). Although compressed capacity wasn't tested, there was plenty of capacity for all tests conducted in this review.

Logger includes a useful analysis portal, complete with dashboards, reports, fast search capabilities, and settings for real-time alerts. Logger leverages the Common Event Format (CEF) for event classification and reporting. Prebuilt reports (284 at time of test) come with scheduling interfaces, access controls, a variety of export formats, and other flexible features.

These and other details are in the test report that follows. This report also provides a requirements checklist for those considering options and features for their log management systems.

[2] www.sans.org/reading_room/analysts_program/LogMgt_June08.pdf

Getting Started: A Requirements Checklist

Some important, up-front advice to getting started is to "try before you buy," particularly when it comes to load testing. Some systems that work well with one or two servers may become unacceptably slow (or drop events) when subjected to log data drawn from a mix of hundreds of routers, switches, and other equipment of various brands and versions. The scalability is also important during security or network events, when capacity and flexibility are most needed. As our 2008 survey shows, businesses will likely want their log management and reporting systems to do things the system developers never thought of as user groups discover more and different benefits their log data can provide.

Here are some other words of wisdom:

1. Deployment and Management

Making the determination between appliance- and software-based log managers depends on the size of the organization and other factors. For medium and small enterprises lacking the manpower and hardware to install a software-based log manager, appliances that also handle log data storage make the most sense. For larger, more complex organizations with resources and staff, appliances or software-based log management that can leverage their investments in storage area network (SAN) storage might be more appropriate.

Also to be determined is whether to use agent versus agent-less collection of logs. There are lots of devices, including some operating in Windows-based systems, that can't natively push their logs to a central location, so many log management systems require agents on these devices for log collection. Some agent-based technologies work well, such as the Snare agent that works on Windows servers and imposes minimal load. However, maintenance and help desk overhead for such point-by-point agents can be problematic for organizations lacking resources, in which case agent-less log management systems make more sense.

Web-enabled administration is also helpful because of its flexibility.

2. Collection

Fifty one (51) percent of respondents to the SANS survey reported dissatisfaction with the most elemental (and critical) step in log management—collecting logs from hundreds of log management formats represented by their platforms and systems.

Breadth of coverage on logging devices, then, becomes a critical log management requirement. The system should support a range of device types from Cisco IOS and Windows operating system logs to home-grown applications and legacy systems. It should be flexible enough to integrate into more systems as needs dictate.

3. Storage

Given regulatory requirements for data retention, log storage can quickly reach ranges in the multi-terabytes. Log management systems or appliances should meet requirements for bundled and efficient storage and/or be able to leverage existing investments in a SAN or network attached storage (NAS) as the primary data store (and/or for backups and archives). How data is stored (in a secondary relational database or within the log management system) is also important to consider when it comes to accessibility.

Support for both raw and normalized data formats should also be considered. While normalized data is faster and easier to parse and access, many IT professionals prefer having the raw data available for review, forensics, and litigation. For some companies, this is an either/or situation. Ideally, however, organizations should have both types of data available, which means log management systems should have the capacity to archive raw data and normalized data. Both types of data stores should also be searchable.

When it comes to retention time, the system should be able to enforce policy based on the device type and duration (as mandated by specific regulations), with the ability to change storage duration requirements as needed. The log storage architecture should support high collection rates as well as high search rates without compromising storage efficiency, and search results should be readily available even when the log management system is under a heavier load than normal.

4. Analysis

Reporting capabilities should come with flexibility in terms of templates and formats (grid and graphical) as well as scheduling and export options. The ability to support top-down, role-based analysis is also helpful for following up on interesting results within reports and for root cause analysis. Another handy feature is the ability to convert new reports into alerts that can be used to detect similar incidents in the future.

Applying a common format across log data sources is also important, as it enables users to easily navigate through log data without being familiar with each source in the enterprise and its specific log syntax. Reporting features should be intuitive to follow and include common security, compliance, and operational content-based reporting features. Users should also have the ability to create new reports without too much trouble.

5. Scalability

Plan for scalability to handle events and any critical use the enterprise requires of its log data. For example, if being able to generate reports to prove PCI compliance is a primary need, then the log management system should align with those reporting needs. If the IT group wants the ability to quickly search through large amounts of log data to find details about a specific incident and store the results for forensics or other purposes, archival storage, retrieval, and incident reporting are strong requirements.

When considering scale, think larger rather than smaller. Processing speeds, search speeds, and storage capacity should be able to handle increases in load—especially when there is an event, such as an internal outbreak that's creating voluminous log data—because this is when system reliability and speed are most needed.

Review: ArcSight Logger 7100 v.3.0

This test of the ArcSight Logger 7100 v3.0 included several areas: setup, collection, storage, and analysis (searching/reporting). Setup was straightforward, and Logger successfully collected log data from all sources in the test network. It was easily searchable for specific data strings and included a variety of report types, although using the reports required some learning of Logger's terminology around log data sources.

Appliance Setup

Thirty percent of the SANS 2008 Log Management Survey respondents indicated that "The Whole Lifecycle" of log management was a critical problem for their organizations, with one comment saying simply, "Getting started."

Logger includes a one-page "Getting Started" instruction sheet that begins with connecting a keyboard, mouse, and monitor, and includes options to use a dumb terminal or Web interface. Once the Logger 7100 booted up, I logged into the appliance and specified key parameters including IP address, subnet mask, gateway, and hostname. On reboot, Logger was ready for configuration.

During configuration, it's important to be aware that capacity should support retention policies and groups. Also, a reliable time source is critical in the analysis of log data. The best option is an internal timeserver. If one is not available, the same external timeserver used by other systems on the network will suffice.

The final setup step is the specification of fields to be indexed. In most cases, accepting the default setup (the setting used for this test) is the easiest option for getting started.

Collection

Fifty percent of our survey respondents cited log data collection as their most critical issue. ArcSight Logger 7100 version 3.0 was designed for collecting events from multiple sources at extremely high rates. Specific collection details include:

Supported Data Sources: ArcSight Logger can collect logs directly from nearly any syslog or file-based log source. It comes with SmartConnectors to collect data from over 275 different products. For legacy sources, ArcSight offers a wizard-driven interface called a FlexConnector to develop custom connectors.

In general, Connectors accept raw log data, normalize it, and forward Common Event Format (CEF) compliant data. This CEF-compliant data can then be indexed by the Logger for high-performance search and reporting. Connectors accept log data using a variety of methods, including a TCP or UDP syslog stream, pulling the logs from Windows servers, and retrieving log files over FTP and other file transfer methods.

The Connectors used in this test were the ones built into ArcSight's smallest Connector Appliance—the ArcSight C1000—which didn't require much technical configuration. This appliance, which comes with Connector software preinstalled, automatically detected my test log data and sent it on to the Logger appliance. In this scenario, a Cisco router, a PIX firewall, a couple switches, and two Windows domain controllers were some of the devices that were configured to forward their log data to the Connector, which correctly interpreted each format and sent the data on to Logger in Common Event Format (CEF). The Connector forwarded both the normalized data and the raw log for forensics purposes.

ArcSight's 275 Connectors analyze the data and convert it to a Common Event Format (CEF). CEF is the industry standard for the interoperability of data generated by logging devices. Some of the comments in our 2008 survey pointed to the difficulty in understanding log data. Part of that difficulty is that every device formats the output a little bit differently. ArcSight Connectors convert the incoming data to the CEF format. The Connectors also have an option to maintain the raw syslog format as one of the fields in the CEF data.

Definition: Common Event Format

The Common Event Format (CEF) is supported by a growing number of companies that generate and process log data. The basic format is:

CEF:Version|Device Vendor|Device Product|Device Version|Signature ID|Name|Severity|Extension

The extension field is the location for the data that the application should be reporting. Once you understand the basic format, simple scripts can be modified to log data in a CEF-compliant manner and the data can be indexed.

Collection Performance: The L7100 officially supports up to 100,000 events per second and was able to consistently store events at considerably higher rates without losing any events when receiving log data over TCP. It also ran log searches and reports while collecting log data at these rates. This makes sense because it's important to have some excess collection performance room to accommodate event spikes, such as would occur during a virus outbreak.

When testing UDP syslog data, some events were dropped. However, this test against UDP syslog data didn't include ArcSight Connectors, which compress, normalize, and forward the logs to the Logger appliance. Events that were sent using Connectors over the more reliable TCP protocol were not dropped.

Device Identification: In the lab, I set up the IP address of 10.1.1.241 for the Logger and added the line, "logging 10.1.1.241," to the Cisco routers and switches, which started sending log data to the Logger using UDP. The Logger immediately recognized the syslog traffic and added the device to a list of logging devices that it had received data from.

Distributed Collection: The C1000 Connector Appliance that was tested is intended for small branch offices or retail locations. Connectors can also be used at regional outlets, such as bank branches, hospitals, retail locations, and all forms of business subnets. It handles local storage, encryption, hashing, bandwidth controls, batching, and caching for uploading data to the central Logger appliance.

Storage

ArcSight Logger 7100 comes with 4.5 terabytes of RAID-enabled physical onboard storage, so I had plenty of space to work with. After accounting for system files and compression, this translates to a claimed capacity of 35 TB, which is substantial for a 2U (mid-sized) log management appliance.

Logs from different devices can be assigned to distinct retention policies. For example, all logs from devices subject to PCI compliance can be assigned to a specific storage group with a configurable retention policy. Similarly, other sources can be subject to a distinct policy depending on corporate or regulatory requirements. I didn't run a full test on this feature. I did, however, run a seven-day test to clear out various types of data (Cisco, Pix, Linux, Windows, etc.) by configuring a storage group with a maximum age of seven days and then verifying that logs were no longer accessible thereafter.

In addition to backup and restore capabilities for configuration data, the appliance can also archive events to any NFS (Network File System) file mount on a daily basis. This worked as expected and is a valuable feature to ensure compliance with regulatory retention requirements. What's nice is that the daily archives do not have to be restored for access and analysis. You can simply link to any number of daily archives and then search and report across your onboard data as well as your linked daily archives. Although I did not test SAN support, ArcSight offers a dedicated model (L5100-SAN) that is equipped with an HBA (Host Bus Adapter) interface and is optimized for use with SANs.

Analysis: Dashboards and Reports

Searching and reporting were problematic for more than 40 percent of respondents to the SANS survey. With Logger, end users are launched directly into personalized dashboards that present multiple related reports in a single view. From these dashboards, users can drill into and across specific reports. Logger came with 88 foundation reports available out of the box, with add-on, for-fee packages of 67 PCI reports and 129 SOX-related reports also available.

The ArcSight approach is to classify events so that reports can span all types of devices and so that end users don't have to be familiar with the various log syntaxes they may encounter. In tests, a single report on "Users Created" consistently showed new users created in the lab's Windows 2003-based domain and a Windows 2000-based domain. It also showed local users on a PIX as well as users on a Linux system. Similarly, a "Device Configuration Change" report correctly showed changes to routers, firewalls, the Linux system, and both Windows domains. Reports even included the setup and teardown of the VPN connections (Security Association deletions and creations) in the device group VPN.

The built-in reports also make a good starting point for building your own reports. For example, the main firewall for the lab is an iptables firewall. To generate a report including all blocked packets arriving on the outside interface (eth1 in this case), the first step was to create a query either by entering SQL or by using a wizard. The most helpful approach to building queries was to duplicate existing queries and modify them as needed.

Modifying prebuilt reports to incorporate additional device types or building brand new reports took a little time because I had to learn Logger's field set and SQL query syntax, but became easier after I familiarized myself with their terminology. For example, in another test, successful and failed connection attempts to both a Cisco PIX and a Microsoft PPTP VPN server were correctly collected, but the events did not show up in the report on failed or successful access attempts. Comparing the event categorization to the query syntax identified the problem: The report was looking for events categorized by ArcSight as coming from VPN devices, while the PIX events were appropriately tagged as a firewall source. I modified the query to include events classified as a VPN or as a Firewall and the report picked up all the events thereafter.

Once created or modified, reports can be saved and used as the basis for new custom content, exported in a number of formats, and scheduled and emailed to administrators, auditors, and business units.

Analysis: Searches and Alerts

When a report suggests a problem or an anomaly, high-speed interactive access to log data is imperative. Then, once the crisis is over, the system should be able to add the event characteristics to a template and recognize future matches of this event category discovered on the network.

ArcSight Logger allows searching of data using plaintext or regular expressions as well as indexed searches based on fields (for fastest results). Searches were intuitive and included the option to drill down endlessly and in different ways, for example searching for specific data as well as excluding specific data—both of which are equally important. In a worm eradication for a client, for example, I used this technique of elimination to identify and then exclude infected IP addresses from log searches in order to continue searching without all the noise from infected addresses that had already been identified.

Logger can be used to search for the signature indicating an infection and then exclude discovered infected machines by right-clicking on the machines' IP addresses, passing those IP addresses on to the helpdesk for remediation, and then refreshing the screen. While no outbreak was simulated in the lab, searches successfully verified that Logger was getting specific data. In one case, I generated traffic against the lab's firewall while generating 6000 EPS (Events per Second) of other log data. I then looked for all data coming from the firewall and gradually eliminated data until I found the packet with the timestamp and ports to match what I was generating.

Regular expression/plain text searches are the easiest starting point for gathering reports, even though they are slower than indexed field-based searches. Once an operator learns what fields align with the data that is being searched, they can achieve much faster searches through indexed field-based searches. According to ArcSight, a speed improvement of a 100x on indexed searches was accomplished in Version 3.0, but comparison across versions was not in scope for this test.

In another test, connection attempts to port 45454 on the outside interface of the firewall (which the firewall blocked) took 26.4 seconds to find using a regular expression search through 12-hours worth of collected data (2.4 million events). When re-executing the search looking for 45454 in the destinationPort field, it took 1.6 seconds. This translates to a search rate of 1.5 Million events per second. The field-based search also yielded higher accuracy because it showed only the destination ports that matched, whereas the freeform search included partial matches such as a Windows log reporting an event id of 745454, unrelated PIX events, and a few other events with 45454 in the description.

Finally, ArcSight Logger allows search patterns and expressions to be saved, shared, or converted into real-time alerts to trigger notification via SMTP, SNMP, syslog, and also directly within the ArcSight Logger Web console. The alerts also support basic anomaly detection. For example you could choose to be notified only if there are five or more matches of an expression (representing a failed login for example) within a certain time span.

Looking Forward

Moving past the scope of this test, the larger framework that a log management system will tie into should also be a question for organizations to consider. While log management isn't the same as Security Information Event Management (SIEM), the lines are blurring between the two. Log management offers the broadest range of use cases from security to audit and network operations, while the SIEM focus is on more targeted cases built around real-time correlation of logs for security and audit purposes. Integration between the two will likely benefit organizations, but this distinction is important when setting business requirements.

The direction in which log management is going will be the focus of our 2009 Log Management Survey (release date April '09). One thing is clear from the 2008 survey: Demand to derive value from log data will continue to rise, even as systems requiring logging will continue to change (for example, virtual systems and cloud networks). Therefore, log management tool vendors need to continue working with standards, their customers, frameworks, and other tools at their disposal to meet these evolving needs.

About the Author

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

SANS would like to thank this paper's sponsor:

ArcSight