ArcSight™

Whitepaper

ArcSight™ Logger

Extracting Value from Enterprise Log Data

Research 002-103108-02

ArcSight, Inc.
5 Results Way, Cupertino, CA 95014, USA
www.arcsight.com info@arcsight.com

Corporate Headquarters: 1-888-415-ARST
EMEA Headquarters: +44 870 351 6510
Asia Pac Headquarters: 852 2166 8302

Executive Overview

Compliance, forensics, security and IT operations teams have long recognized the value that log data can deliver. An effective log management solution can help organizations in several ways:

- Contain the growing cost of regulatory audits through automation
- Reduce expenditure on point security and compliance tools through comprehensive monitoring across all users and systems
- Cut data center costs through consolidation of siloed homegrown log infrastructure
- Improve efficiency of forensics investigations with high-performance log analysis
- Increase troubleshooting turnaround times and adherence to SLA's

Despite these tangible benefits, organizations continue to struggle with even the basic steps of log management such as collection and analysis. This whitepaper will outline the drivers for log management as well as their underlying challenges and drive towards a common set of requirements for evaluation of log management tools. The paper also provides an overview of the ArcSight log management solution and concludes with several examples that illustrate how enterprises can leverage an effective log management solution to automate security monitoring and regulatory compliance, conduct forensics more efficiently and improve operational standards.

Consumers of Log Data

Across the enterprise, there are a growing number of constituents that can benefit from log data.

- Audit and Compliance Groups recognize the value of log data in monitoring adherence to compliance controls and in simplifying, automating and streamlining costly compliance initiatives. Manual efforts and homegrown log infrastructure may provide a patch solution for initial audits, but do little to deliver long-term cost reductions in the face of extended regulatory data retention and stringent audit reporting requirements. There is a clear need for a comprehensive log management solution that can provide efficient collection and low-cost, long-term storage of audit-quality log data from regulated sources, ranging from networking equipment and security devices to databases and homegrown applications.
- Security Teams can leverage rapid access to log data for security threat detection, investigation follow through and development of remediation plans. To facilitate those benefits, log management solutions need to support analysis of log data over extended periods of time, as well as isolating events based on common attributes such as source type, user name, IP address, etc.
- IT Operations and Helpdesk Teams responsible for networks, security or applications are working more closely together or even merging, and they can certainly benefit from a consolidated view of operational activity across the enterprise. To meet operational objectives around availability and SLAs,

Figure 1: Consumers of Log Data

the complexity of consolidating log information across disparate and functionally-oriented event sources must be addressed. An efficient and scalable log management infrastructure solves this problem by supporting high-volume log collection across all network sources with the added flexibility of simplified analysis and contextual data for improved operations.

- Executives (CIOs, CFOs and CEOs) can benefit from dashboards and reports that provide ongoing visibility into business objectives, operational metrics, corporate governance and regulatory initiatives. Summarized snapshots combined with substantive facts can provide a quick assessment of progress and posture against these goals.

Log Management Challenges

Logs are nothing new and most devices are capable of generating logs natively. So, why are companies still struggling with Log Management? The SANS Institute conducts an annual Log Management survey and their survey regarding Log Management challenges suggests that companies are still struggling with collection and analysis of data more than anything else. In fact, nearly half of all respondents indicated that this was still the most challenging aspect of Log Management.

Roughly another quarter of the respondents indicated that securing storage and establishing chain of custody were also big problems around Log Management and about one third suggested that the entire lifecycle of collection, storage, and analysis was problematic for them.

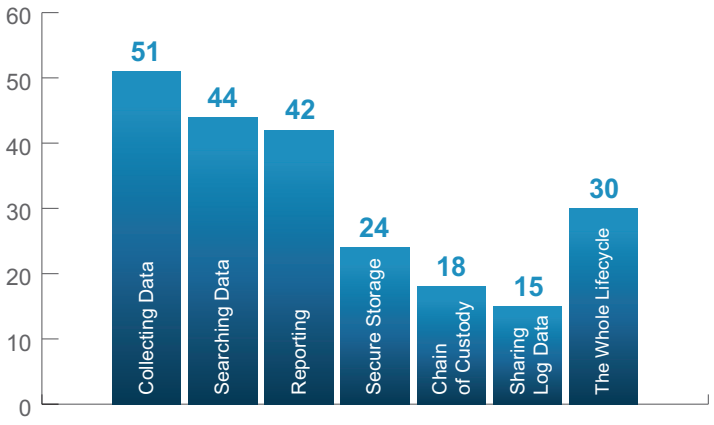

Figure 2: Why are Companies Still Struggling?
SANS Annual Log Management Survey - June 2008

Log Collection Challenges

Log collection is a problem for several reasons, but the scope of collection is perhaps the biggest one. Especially as a result of compliance, organizations have to collect logs from numerous devices and device types all the way from security / network devices up through operating systems, databases, as well as applications and web logs. Simply keeping up with the growing log volumes can be a challenge.

Many of these devices can not natively push their logs to a central location, so companies often deploy log collection agents. If this is not already sounding painful, consider a larger network where multiple locations with no IT staff are involved. Who is going to rollout and manage the log collection infrastructure? What if critical servers do not have any more processing capacity to host log collection agents locally?

There is also the challenge of establishing audit quality when logs are being collected – demonstrating that logs were collected securely and reliably. For log collection from remote locations (stores, bank branches) it is also important to ensure that low bandwidth links are not saturated with log traffic because that can impede the more important transactional traffic.

Log Storage Challenges

Beyond collection, log storage is another big challenge for many companies. Increasingly, organizations are storing more logs and for longer periods. This increase can be attributed to the data retention requirements from the slew of regulations that organizations are now subject to. For example:

- The NIST Guide for HIPAA suggests that logs be maintained for a minimum of six or seven years, depending on the type of data.
- Similarly, section 103 of the Sarbanes-Oxley Act states that enterprises must store all audit work papers and other information related to any audit report, in sufficient detail to support the conclusions reached in reports for a period of seven years. Since it is common to leverage log data to address reporting requirements, many companies interpret the SOX retention requirement as being applicable to relevant log data.
- Section 10.7 of the PCI Data Security Standard explicitly requires governed entities to “retain audit trail history for at least one year, with a minimum of three months online availability.”

Table 1 summarizes data retention requirements for common regulations.

Regulation	Retention Requirement*
SOX	7 years
PCI	1 year
GLBA	6 years
EU DR Directive	2 years
Basel II	7 years
HIPAA	6 or 7 years
NERC	3 years
FISMA	3 years

Table 1: Data retention requirements for regulated data. *Values are regulatory statements and/or ArcSight interpretation of regulatory statements.

When logs are left on their native devices, it is very difficult to enforce retention policies and any type of log rotation becomes a manual, tedious and error-prone process.

Finally, enforcing access controls or establishing the integrity of the stored logs is even more challenging when logs are strewn across devices on your network rather than being consolidated in one location.

Log Analysis Challenges

Analysis remains the third major challenge, especially with homegrown solutions. The fact is that the various different types of devices out there – firewalls, routers, switches, and the numerous different applications and databases out there – all generate logs in a distinct and very often cryptic format that is difficult to analyze. Homegrown solutions provide no simple mechanism to search and report against this information and with the growing community of non-technical consumers of log data there is a definite need for a simple way to analyze logs.

“High-performance” search and reporting is a very valuable and visible part of the end user experience across all log management use cases. For example if an insider threat is detected by a SIEM solution, one of the immediate tasks is to figure out the scope of the impact and that requires quick and easy interactive access to logs to build an evidentiary trail. Similarly, In IT Operations, when critical transaction enabling systems or applications are down, every second can be quantified into lost revenues and the speed of problem identification and resolution is critical.

An additional challenge is that with regulatory mandates, organizations need expertise to build authoritative content in the form of reports, alerts etc. Often such expertise typically does not exist in-house and it is certainly expensive to acquire as a service.

Choosing the Right Log Management Solution

The right solution has to address the challenges outlined in the previous section around collection, storage, and analysis. However, it must also be delivered with simplicity in terms of deployment as well as ongoing management. The right solution also has to offer small- to enterprise-scale or simply put the right scale. Choosing the wrong solution can quickly lead to a log management nightmare, especially as regulations continue to evolve and security threats multiply. So evaluation best practices are a valuable tool. The following list of ten important evaluation criteria has been compiled from the experiences of customers and can help any organization in picking a solution that is right based on their specific use case.

1. Universal Collection

The average enterprise today supports an array of devices ranging from firewalls and IDS/IPS systems to routers, databases, operating systems and applications. Effective log management, especially for regulatory compliance, necessitates collecting event logs from a wide variety of sources across distributed networks. This makes out-of-the-box coverage and device support a critical log management requirement. Make sure the selected solution supports a wide range of device types while also delivering support for all leading vendors.

Log management solutions should deliver raw as well as analysis-optimized log collection from any source. Compliance in particular will justify the inclusion of application logs—both commercial and homegrown. It is paramount for the log management solution to support non-traditional, in-house/legacy applications as well.

2. Performance without Compromise

Across the growing number of use cases for Log Management there is a definite trend towards:

- Including more devices and more logs - which requires high-performance collection
- Longer term retention - that drives the need for more storage efficiency
- More consumers of log data - suggesting the need for faster and simpler log analysis

Most solutions give you the ability to do only one of the above while significantly compromising the others, or require more hardware. Ideally you want performance without such a tradeoff. Fewer appliances will mean lower direct hardware cost but will also reduce power consumption and support for “green” initiatives.

3. Personalized Analysis Portal

Log management solutions must deliver rich analysis capabilities within a personalized, role-based portal. Each user should be presented with interactive and personalized dashboards that combine relevant content into a single, role-based view with access controls.

A top-down view into enterprise-wide events provides a concise starting point for analysis. It also minimizes the need for manual interpretation of results via exporting to spreadsheets. Auditors would certainly prefer an aggregated view of compliance rather than sifting through hundreds of reports.

Reporting capabilities should come with extensive flexibility in terms of report templates, grid and graph formats, as well as export options. To follow up on interesting results within reports, the same analysis portal should enable structured or unstructured searches for rapid root cause analysis. When reports or ad-hoc searches reveal anomalous activity, the logic should be easily convertible into an alert to detect similar future incidents. Alerts not only need to be evaluated continuously against any incoming stream of log data – they must also support flexible logic and notification options.

All content should leverage a common format that allows end users to easily navigate through log data—regardless of their familiarity with source-specific log syntax. This will also eliminate content explosion and the need for device or vendor specific analysis.

4. Bi-directional SIEM Integration

Log management is frequently a stepping stone towards more focused use cases for log data including real-time, cross-device correlation of events for detecting perimeter and insider threats as well as policy violations. It is paramount that your log management infrastructure interoperates seamlessly with your SIEM investment because very often the users are the same and certainly the same underlying data (logs) are the same.

The integration should be bi-directional, allowing the log management solution to forward the relevant subset of events to the SIEM tool for further analysis. In turn, your SIEM tool should be able to send events representing detected threats back to the logging appliance for search, reporting and archival. To enable a chain of custody, communication between the log management and SIM infrastructure should be reliable and secure. Finally, both investments should leverage a common collection infrastructure to avoid unnecessary deployment and maintenance overhead.

5. Efficient and Self-managing Storage

Once the decision is made to capture all logs across the enterprise, the data must be stored efficiently and effectively. Organizations that implement homegrown log infrastructures often end up with silos of distributed log servers that are very hard to manage. Given regulatory requirements for data retention, an organization’s log infrastructure can quickly reach multi-terabytes. Demand a solution that can provide efficient storage through compression without compromising rapid analysis. Log management solutions should also allow you to leverage external SAN storage and any bundled storage should come with built-in RAID protection while also supporting configuration data backups and archival of actual log data.

Finally, retention policies should be automatically enforced based on the device type and duration mandated by specific regulations, with the ability to extend the life of the logs in the event of impending litigation.

6. Audit-quality Logs

The use of logs in compliance audits and litigation requires that organizations be able to demonstrate the integrity and availability of log data in transit and at rest. Even a few missing or compromised log events can lead to inaccurate audit results or may create doubt around the validity of audit reports. To easily demonstrate, audit-quality collection has to occur close to the event generating sources, which highlights the importance of distributed collection. Log collection infrastructure in remote locations can provide

buffers to prevent data loss when network connectivity to the data center is lost. It can also enable a secure and reliable conduit to transfer the log data back to the centralized log repository. This safeguards data as it travels through the network, ensuring there is no loss or alteration of logs.

Also look for the ability to preserve logs in their original form. The ability to prove that captured logs have not been tampered with or modified is another key audit-quality best practice. Integrity checks address that requirement but make sure a robust hashing algorithm sanctioned by the NIST 800-92 Log Management standard is being used. Finally, high-availability measures such as failover capabilities from one data center to another can further minimize data loss.

7. Ease of Deployment and Management

For some large enterprises, form factor flexibility may be of value but most organizations today look for rapid deployment and want to minimize add on services – and in that context appliances have several advantages:

- Avoid delays and hassles associated with selecting, testing, procuring, and deploying the hardware separately – particularly important if you have limited staff overall or in remote locations such as stores
- Less management overhead – hardware, software, and OS updates are all managed by the same vendor
- Bundled storage is usually included with appliances – whereas software solutions require a separate procurement cycle and storage management expertise

If there is a definite preference for a software solution, at a minimum do not forget to account for the incremental cost of acquiring, setting up and managing separate hardware and log storage.

In terms of ongoing management, agentless collection is a big benefit. There are lots of devices, including Windows, that can not natively push their logs to a central location, so many solutions require agents. But who is going to rollout and manage the collection agents? And more importantly what if critical hosts do not have any more processing capacity to host log collection agents locally? Whenever possible, an agentless yet distributed collection option is always going to be easier.

Finally, logs should be stored in a compressed format that does not require database administration expertise. Web-based access for end users will further simplify deployment by eliminating the need for client installations and allowing you to easily scale end-user access.

8. Distributed Event Collection

An architecture that supports cost-effective, distributed event capture is essential to ensure complete collection of all logs without data loss in distributed environments like banks and retail environments.

Collection of logs from remote locations (stores, retail sites, branch offices, etc.) requires additional bandwidth, which is typically limited or already saturated between remote locations and data centers. This constraint has multiple repercussions. First, it is important to ensure that mission-critical transactional traffic is always prioritized over all other traffic. Log management solutions must provide controls to limit the use of bandwidth for logs. Additional optimization measures, such as compression of logs in transit, can also mitigate the impact of saturated or low-bandwidth links.

A more sophisticated approach would combine bandwidth controls, compression, and prioritization with effective batching techniques by time-of-day and by event severity. This can ensure that while limited bandwidth is used for collection of important security-relevant logs, other logs can be batched to the data centers for storage and analysis during off-peak hours when a store is not in operation or more generally when there is more bandwidth available.

But most organizations know this! In other words they recognize the importance of distributed log collection. The big obstacle has always been the absence of a cost-effective distributed log collection option. So the requirement here is actually not distributed event collection but rather “cost-effective” or feasible-distributed event collection. Simply put, ask how much it will cost you to place a collection appliance in each remote location to ensure audit quality.

9. Small to Enterprise Scale

Companies often begin their search for a commercial log management solution with a given driver in mind, such as SOX compliance. Over time coverage is expanded to other regulations as well as other consumers of log data such as security, operations and networking. As the number of sources increases, so does the overall event rate and event volume. This trend highlights the importance of scalability in log management evaluations. You want a log management solution that leaves room for short-term growth, but one that can just as easily grow as the scope of your Log Management effort evolves. It should not require an unnecessary upfront investment in excess storage or event rate capacity. With a one-size-fits-all solution, you typically get a system that, at first, offers more than you need, but then fails to scale to your long-term needs.

Log management solutions must support the ability to easily add more sources and device types. Make sure a range of performance options are available so you can expand coverage without more hardware. Similarly, the centralized log repository should also be offered in a range of performance (EPS) and capacity options so that you can scale storage and performance linearly. Regardless of whether deployment is hierarchical or peer-to-peer, your log storage infrastructure should always provide a universal view into corporate-wide log data.

10. Pre-packaged Content

Common security and operational reports should be available out-of-the-box as part of system content. A solution that expects you to build all the content is more of a platform than a real solution. There is a definite cost associated with building content as well, so factor in associated delays if the content you need is missing.

Regulatory compliance mandates impose extensive audit-reporting requirements, and while some log management solutions provide packaged regulatory content, few demonstrate how the content maps to authoritative sources. Make sure packaged content is available to automate audit efforts. Add-on content packages for specific regulations should derive from best practices and authoritative standards like NIST 800-53 and ISO 17799.

When you make changes or build new content for your use case, portability of that content across Log Management analysis appliances or more generally instances is also key. You do not want to rebuild content everywhere. Ideally you want to build content once and roll it out to all locations easily – and that is why portability is key.

ArcSight Logger

To address the growing need for collection, storage and analysis of enterprise-wide log data, ArcSight Logger is offered in a range of turnkey, stackable appliances that support high-performance collection of logs from any source into a highly-compressed yet accessible and self-managing log data repository. With a powerful reporting and alerting engine, ArcSight Logger functions both as a standalone appliance for log management as well as a complement to the broader portfolio of ArcSight products. The components of the ArcSight Log Management solution include:

- **ArcSight Connectors** - which provide comprehensive collection of all logs within your network with audit quality
- **ArcSight Logger** - which supports fast collection, efficient storage, as well as quick and easy analysis of logs
- **ArcSight Solutions** - deliver pre-packaged content for specific-use cases such as PCI or Sarbanes-Oxley compliance

The following success stories highlight the capabilities of the ArcSight platform through actual deployments and customer success stories.

Real-Life Scenarios – Log Data Completes the Picture

Enterprises are leveraging the capabilities of the ArcSight log management solution —both within and beyond the enterprise's security organization. The following sections will highlight some of the traditional use cases for log data and reveal some new use scenarios that ArcSight enables.

Example One: PCI Compliance

A nationwide apparel retailer faced a major network breach leading to the theft of credit card information for millions of its customers. As a result, the company experienced brand damage, loss of customer trust as well as lawsuits and fines.

As a Tier 1 (VISA classification) merchant, the organization was already subject to the PCI DSS standard, but had not yet implemented an enterprise-wide log management solution (explicitly required under Requirement 10 of PCI), nor the broader monitoring requirements mandated by PCI.

In response to this breach, the organization embarked on a search for a log management solution to automate PCI audits for cost reduction, and proactively protect their cardholder network and infrastructure for real-time visibility and response to threats. However, with 1000s of retail locations linked to regional data centers via low bandwidth links, audit-quality collection from remote locations and centralized storage/analysis of logs, while critical, would be difficult to accomplish without impacting transactional data traversing the same links.

After a thorough evaluation, the retailer chose ArcSight based on the following success criteria:

- Universal event collection: ArcSight provides the ability to flexibly pull in logs from widely-prevalent homegrown and legacy systems at the retail locations

- Cost-effective store level log collection: ArcSight is uniquely able to deliver low-cost turnkey appliances for localized collection at the store level. This architecture enables audit-quality collection by providing secure and reliable transport as well as buffers in the event of network outages. It also addresses their transaction assurance requirements through a combination of bandwidth controls, batching, prioritization and compression.
- Powerful analysis: ArcSight Logger provides a powerful and personalized role-based portal with interactive dashboards, drill-down reporting, rapid ad-hoc search and intelligent alerting capabilities. Only ArcSight Logger could deliver search rates in the range of 3,000,000 events per second without imposing a significant tradeoff on collection rates or storage efficiency.
- Comprehensive out-of-the-box content for PCI: Pre-packaged PCI content enabled this organization to kick-start their compliance initiatives without requiring extensive training or services. Reports, alerts and dashboards, as well as real-time PCI threat detection correlation rules were clearly mapped to the actual PCI audit requirements.
- Integrated log management and SIEM platform: ArcSight Logger is tightly integrated with its market-leading SIEM solution (ArcSight ESM) for seamless interoperability across the historical log repository and the real-time PCI monitoring solution.

As a result of its ArcSight implementation, this nationwide retailer has been able to automate the majority of its PCI audit-reporting requirements, which has yielded significant savings relative to their homegrown log management initiative. More importantly, the retailer now has real-time visibility into potential threats, which enables rapid response and timely containment of risk.

Example Two: Intellectual Property Leak Investigation

A key R&D employee had been recruited by a competitor and left the organization. There was concern among the terminated employee's peers that proprietary and confidential information may have been accessed and siphoned away by this employee. A request was made by senior management to get visibility into all systems and data that the terminated employee accessed.

At first glance, this appeared to be as simple as scanning logs to determine what files were downloaded and what hosts and applications were accessed by the terminated employee. However, it soon became clear that all kinds of network activity needed to be analyzed to determine if the terminated person changed configurations, accessed confidential intellectual property, planted hidden executables, opened a backdoor to the network or did something else not associated with his normal level of access.

In investigations like this one, the request will typically begin as a "show me all activity conducted by the terminated employee during the last month prior to termination." The next level of analysis could require reviewing the employee's activity over a longer period to look for trends of unauthorized access, higher than usual printing activity during off hours and other anomalous activity. These processes can be both difficult and time-consuming for an organization with a large volume of event logs generated across data centers and consolidated at many discrete locations—often the case in homegrown systems. An employee can generate a lot of activity within a short timeframe, and that data would most likely be distributed widely throughout the enterprise's IT network and various system components.

With ArcSight Logger, the enterprise was able to efficiently collect, centrally store and provide interactive, search and reporting capabilities across all log data. The analysis spanned multiple ArcSight Logger appliances distributed across the organization.

This particular request required only a simple, one-step term search on potential usernames and identities the employee may have had access to. Results were then presented as exportable, audit-quality data that was drilled into for further analysis. Whether IT needs to go back two weeks or even two years – ArcSight Logger returns results in seconds.

Example Three: IT Infrastructure Monitoring and Troubleshooting

A portion of the network that serves the marketing, sales and finance department of a busy enterprise has gone down. There had been intermittent problems with this network segment over the last two months.

A variety of network and system management tools, such as network monitoring, performance monitoring and traffic analysis tools were currently in use. While these tools provided real-time alerts of the particular incidents and detailed traffic information at those particular times, they lacked overall system, logical and time-based context—referred to as contextual data. In this case, the intermittent problem meant the networking team had to manually access many machines, devices and systems to determine which logs to examine.

With ArcSight Logger, the investigation team was able to rapidly sift through all log data enterprise-wide to look for clues. Using the drill-down capabilities of ArcSight Logger, the investigator quickly segmented the log data by time and device.

Using this intuitive interface to interact with data allowed the investigator to pose “what if I look for this pattern” type questions very quickly. All search parameters were highlighted for further filtering. The IT team began their search based on the IP address that reported problems, the hostname of the application they were accessing and the application they were running.

By dynamically modifying the search timeframes, they discovered the problems occurred during heavy application access to a set of particular servers. Without the flexibility of analysis offered by the ArcSight solution, the IT team would have spent many hours manually accessing individual logs and sifting through them—and in the interim the cost of the network outage would have continued to compound itself.

In this example, ArcSight Logger provided a valuable component to the logical troubleshooting process that IT was required to go through each and every day. ArcSight Logger allowed the IT team to drill down and intuitively navigate through the information so that root-cause analysis was fast and simple. With this capability, ArcSight Logger improved IT service levels by dramatically simplifying troubleshooting for the helpdesk.

About ArcSight:
ArcSight (NASDAQ: ARST) is a leading global provider of compliance and security management solutions that protect enterprises and government agencies. ArcSight helps customers comply with corporate and regulatory policy, safeguard their assets and processes, and control risk. The ArcSight platform collects and correlates user activity and event data across the enterprise so that businesses can rapidly identify, prioritize, and respond to compliance violations, policy breaches, cybersecurity attacks, and insider threats. For more information, visit www.arcsight.com.

ArcSight™

To learn more, contact ArcSight at: info@arcsight.com or 1-888-415-ARST