RSA

The Security Division of EMC

White paper

Building an Infrastructure That Enables Log Management Best Practices

A Technology Strategy for Comprehensive Security Information and Event Management

Executive Summary

The current regulatory environment and threat landscape make it essential that organizations worldwide become much more strategic about log (event) management. To help organizations develop a log (event) management capability, RSA has developed a series of white papers, which include a set of recommended best practices.

The first paper in the series provides the rationale and methodology for developing best practices as well as covers the details of RSA's set of 40 recommended best practices. This, the second paper in the series, takes the next logical step by guiding organizations in establishing the criteria for an infrastructure to help realize best practices and build a technology strategy for comprehensive security information and event management.

Given the demands for accurately collecting, reliably reviewing, and securely storing logs from across the enterprise – and doing it all in a consistent and cost-efficient manner – the only practical approach to realizing best practices in log management is to build a centrally-managed, dedicated infrastructure.

Since the volume of log data is constantly increasing and retention periods are becoming longer, the centrally-managed, dedicated infrastructure must achieve an information lifecycle management (ILM) strategy for log data. Through an ILM strategy, the organization aligns the business value and/or use model of the log data with the most appropriate and cost-effective storage mechanism.

Building an infrastructure to incorporate an ILM strategy for log data involves combining security information and event management (SIEM) technology with tiered storage. Although infrastructure requirements will vary depending on each organization's particular environment, there are many that are common across organizations. This paper is intended to help organizations determine their own infrastructure requirements by providing a comprehensive look at general requirements and specific requirements in log generation and capture; log retention and storage; log analysis; and log security and protection.

Contents

Definition of Log (Event) Management

This white paper is the second in a series of three white papers on log (event) management. In each paper, the definition of log (event) management is provided in order to clarify the meanings of terms used throughout the series.

A log is a record of an event or activity occurring within an organization's systems or networks. Examples include a firewall allowing or denying access to a network resource, a change to the configuration of the operating system performed by an administrator, a system shut-down or start-up, a user logging in to an application, or an application allowing or denying access to a file. For more examples of events or activities, please see the companion white paper "Log Management Best Practices: The Foundation for Comprehensive Security Information and Event Management," Appendix 1, "Sources and contents of logs."

Log (event) management is the collection, analysis (real-time or historical), storage, and management of logs from a range of sources across the enterprise including security systems, networking devices, storage systems, operating systems, and applications.

Log management is the foundation for comprehensive security information and event management (SIEM) including the following use cases:

- Real-time threat detection and mitigation
- Incident investigation and forensics
- Compliance to regulations and standards
- Capacity planning, performance and uptime
- Evidence for legal and human resources cases
- Detecting and preventing IP theft
- Auditing and enforcing employee productivity
- Troubleshooting system and network problems
- Auditing and enforcing IT security policy

Becoming More Strategic About Log Management

The current regulatory environment and threat landscape make it essential that organizations worldwide become much more strategic about log management. Logs provide a way to monitor systems and keep a record of security events, information access and user activities.

Gone are the days when organizations could disregard log reviews or purge logs with little to no planning. There are now legal obligations and compelling business reasons for an organization-wide, well-orchestrated effort to ensure comprehensive collection, timely review and ample retention of logs.

While it is true that over the past few years many organizations have at least started to focus on log management, the emphasis has now shifted. Log management is no longer just about correlation analysis in order to detect threats in real-time and reduce false positives; or about generating the right reports to meet compliance requirements, although these are both important and becoming more critical.

Because of the increasingly massive amounts of logged data and longer-term retention requirements, organizations now have to also look at log management from the perspective of handling large volumes of log data over its entire information lifecycle. The need to maximize the value, while minimizing the costs of generating, analyzing, storing, retrieving and ultimately disposing of log data has come to the fore. Throughout the lifecycle of log data, a range of stakeholders – from security operations and forensic investigators to the legal department and auditors – must have ready access to this information.

As such, developing best practices in log management is an essential component of an enlightened IT strategy. By establishing best practices in log management, information executives can deliver tremendous value to an enterprise or government agency by avoiding costs and increasing efficiencies in areas such as compliance, risk management, legal, forensics, information storage and security operations.

RSA has developed a set of 40 recommended best practices covering logging policies, procedures and technology; log generation and capture; log retention and storage; log

RSA Recommended Best Practices*

I. Logging Policies, Procedures and Technology (LP)

LP1. Provide executive management support for the policies

LP2. Fully document the policies, perform periodic reviews and update as needed

LP3. Define roles and responsibilities and provide proper support for staff

LP4. Align log management policies with related policies and procedures

LP5. Ensure separation of duties

LP6. Establish standard log management operational procedures to support and maintain the policies

LP7. Plan and implement a dedicated log management infrastructure to support and maintain the policies

LP8. Don't go it alone; leverage an industry expert and/or consultant

II. Log Generation and Capture (LG)

LG1. Ensure logging is enabled on all systems, devices, and applications

LG2. Don't try to filter the logs at the source

LG3. Ensure that data captured includes all of the key events and activities required by the standards

LG4. Ensure that details captured for events and activities are in accordance with the standards

LG5. Ensure that you are able to track individual users by capturing unique user identification information for all user actions

LG6. Test your logging facility

LG7. Synchronize time stamps

LG8. Watch for sensitive data collection

LG9. Consider privacy issues when setting up user activity monitoring

III. Log Retention and Storage (LR)

LR1. Retain logs in a secure, well-managed storage infrastructure

LR2. Use an information lifecycle management approach whereby logged data is stored relative to access requirements

LR3. Retain production data on-line for one year and one quarter = 15 months (min)

LR4. Retain back-up data near-line for the same period as production data

LR5. Retain active archive data for approximately 2-7 yrs (min) using near-line storage for more recent records (e.g. up to 5 yrs) then possibly moving some records to off-line storage (e.g. 5 yrs +)

LR6. Ensure log retention policies are developed in consultation with legal counsel

LR7. Address the preservation of original logs

LR8. Ensure retired logs are disposed of

IV. Log Analysis (LA)

LA1. Regularly review and analyze logs

LA2. Aggregate logs using a centrally-managed log management infrastructure

LA3. Automate as much of the log analysis process as possible

LA4. Leverage correlation tools to attain a holistic view and reduce false positives

LA5. Use automated reporting tools to facilitate review of logs through report generation

LA6. Review procedures should include real-time monitoring of applicable log events

LA7. Set up an alerting system based on priorities

LA8. Develop a baseline of typical log entries in order to detect unusual or anomalous events or activities

V. Log Security and Protection (LS)

LS1. Secure the processes that generate the log entries

LS2. Limit access to log files

LS3. Implement secure mechanisms for transferring log data

LS4. Protect log files in storage

LS5. Protect the confidentiality and integrity of log files

LS6. Provide adequate physical protection for logging mechanisms and stored logs

LS7. Maintain business continuity for logging services

*Excerpt from "Log Management Best Practices: The Foundation for Comprehensive Security Information and Event Management" white paper from RSA, The Security Division of EMC

analysis; and log security and protection. (For more information, please see the companion white paper entitled, "Log Management Best Practices: The Foundation for Comprehensive Security Information and Event Management.") A list summarizing these best practices can be found on page 2.

RSA Recommended Best Practices address:

1. The need for more effective threat detection to mitigate the increasing risk of intrusions
2. The higher levels of audit and reporting demanded by regulations and relevant standards
3. Forensic requirements in light of today's more sophisticated, targeted attacks, which can take place over long periods of time
4. Legal protections such as retaining the right evidence in case of litigation
5. Business objectives such as increasing the productivity of security operations and reducing the costs of compliance audits.

RSA Recommended Best Practices are laid out in detail in the first white paper of this series. They are intended to help organizations develop their own complete set of best practices. This second paper takes the next logical step and guides organizations in establishing the criteria for a log management infrastructure to help realize best practices and build a technology strategy for comprehensive security information and event management. This includes a focus on achieving an information lifecycle management strategy for log data.

Using a Centrally-Managed, Dedicated Infrastructure

An infrastructure for log management consists of the hardware, software, networks, and media used to generate, secure, transmit, store, analyze, and dispose of log or event data. In many organizations today, the handling of logs or events has grown organically. As a result, it is often performed in a fragmented way in which individual departments or stakeholders deploy their own solutions dedicated solely to the logs or events for their particular systems, networks or applications; without any overall guiding strategy. Yet given the demands for accurately collecting, reliably reviewing and securely storing logs from across the entire enterprise, a centrally-managed, dedicated infrastructure is the only practical approach to realizing best practices in log management.

Otherwise, if the alternative fragmented or "siloed" approach is taken, it is difficult or even impossible to implement consistent policies and procedures across the organization and to get a complete and coherent picture of an organization's systems and networks. There will likely be gaps in policy enforcement since administrators only see what is happening on individual systems. Furthermore, attacks may not even be recognized because an anomaly on one system viewed in isolation may seem innocuous.

Whether your organization is a small to mid-sized business (SMB) or a large enterprise, a centrally-managed approach will also be more cost-efficient. For large global organizations, this kind of infrastructure can be implemented using a distributed deployment.

A centrally-managed approach makes it easier to implement best practices in every category of log management. For example, it facilitates oversight of the collection activity across the organization. If log collection is done on a system-by-system basis, administrators will often consider logging a non-priority and turn it off. It improves analysis: by aggregating logs, events can be correlated and displayed via a central monitoring station to give a more informed and holistic view of network events. Adequate security controls can also more easily be put into place and maintained across-the-board.

With a centrally-managed approach, it is also easier to achieve consistent and reliable log data storage. Otherwise, it is difficult to establish and enforce log retention policies: logs with little value from some sources may be getting stored while important records from other sources may not be stored at all. As well, if individual departments or stakeholders deploy their own storage, it becomes costly and complex to manage. A siloed approach results in inappropriate resource deployment, misused storage capacity, degraded performance, decreased availability of log information, and premature investments in yet more new storage subsystems.

In terms of the type of storage resources, a centrally-managed approach best aligns with the use of shared storage versus dedicated storage. Often this means using networked storage systems. Networked storage systems provide for retention and retrieval of log data over a network, allowing users across the organization to have

secure, role-based access to the shared storage devices containing the log data. Unlike directly-attached storage (e.g., a storage system dedicated to a server), networked storage systems deliver economies of scale through device consolidation, resulting in more storage capacity across the organization at a lower cost. Networked storage systems also facilitate centralized administration of storage devices and support heterogeneous computing environments (e.g., Windows and UNIX systems).

Laying the Foundation for Comprehensive Security Information and Event Management (SIEM)

As mentioned in the first section, log (event) management is the foundation for comprehensive security information and event management (SIEM). At one time, many solutions were marketed either as security event management (SEM) systems, which focused on incident response applications, or security information management (SIM) systems, which focused on auditing applications. Today, an infrastructure for log (event) management should address both of these broad categories of applications: compliance and real-time monitoring. SIEM technology provides a platform for aggregating log and event data from across the organization for management, retention and analysis.

Many organizations begin with a need to improve log management to meet compliance and audit requirements but quickly realize that they have many other needs which can be met by using log or event data, such as detecting external or internal attacks and improving responsiveness. Log or event data is being recognized as a wealth of intelligence information. Pooling this information to be handled by a single infrastructure enables organizations to maximize the value of the data.

An optimal infrastructure should perform comprehensive log collection and management for the full range of use cases including: real-time threat detection and mitigation; incident investigation and forensics; compliance to regulations and standards; capacity planning, performance and uptime; evidence for legal and human resources cases; detecting and preventing IP theft; auditing and enforcing employee productivity; troubleshooting system and network problems; and auditing and enforcing IT security policy.

Building an Infrastructure to Achieve an Information Lifecycle Management (ILM) Strategy

The increased volume of log data and longer retention periods create the need for the development of an information lifecycle management (ILM) strategy for log data. Through an ILM strategy, the organization aligns the business value and/or use model of the log data with the most appropriate and cost-effective storage mechanism. By looking at how the value and use of the information will change over time, organizations are empowered to deploy storage resources more effectively.

By combining security information and event management (SIEM) technology with tiered storage, organizations can build an infrastructure incorporating an ILM strategy. The marriage of SIEM and tiered storage matches the need for a centrally-managed, dedicated infrastructure, which can automate log and event management functions throughout the lifecycle of log (event) data and reduce the costs of security and compliance programs.

Tiered storage of log data also enables the implementation of best practices in log management. According to RSA Recommended Best Practices, log data should be stored based on the access requirements of the information, which include the stage of the data as described below.

The stages of log data include:

- Production data, which is being actively used for real-time analysis, on-going review, and periodic audits and assessments
- Backup data, which is a mirror image of the production data that may be needed in case the production data is compromised or damaged
- Active archive data, which is a sub-set of the production data that will be stored longer-term for record-keeping purposes based on regulatory, legal discovery and possible forensic requirements

With tiered storage, data that requires frequent or ready access, such as production data, is stored on-line while data not requiring as frequent or ready access, such as back-up and active archive data, is stored near-line or off-line. Well-designed tiered storage will optimize the use of storage resources, matching the required accessibility and the necessary capacity to the stage and age of the data. It

will also allow secure role-based access to the log data for all stakeholders and secure deletion of the data when it is no longer required.

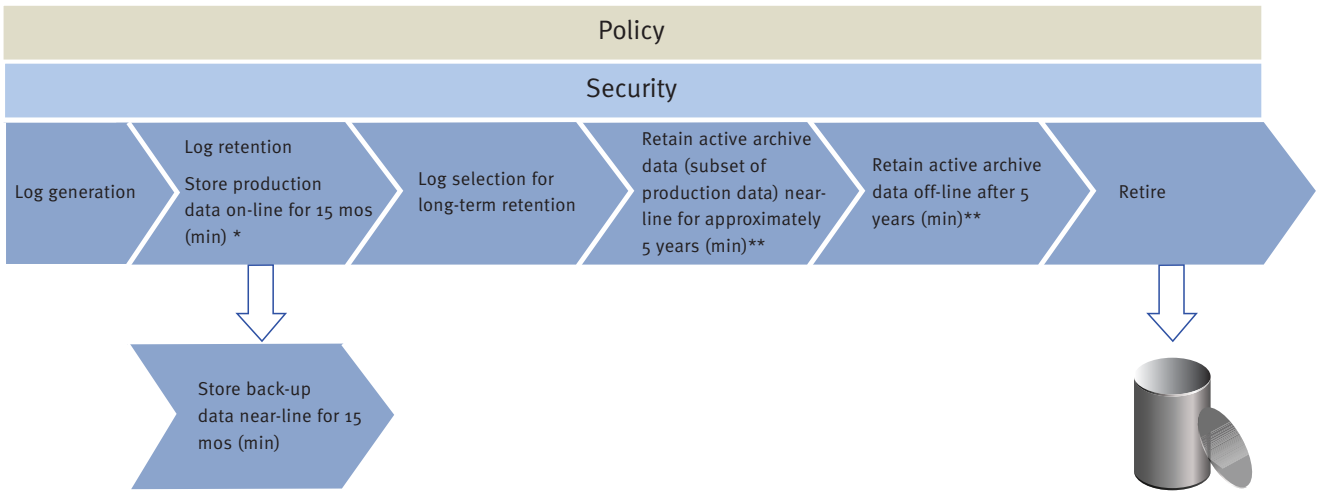
As part of an ILM strategy for log data, an organization will need to determine its retention period for production, back-up and active archive data based on its own policies. RSA Recommended Best Practices suggest that an appropriate retention period for production data is a minimum of one year plus one quarter; for back-up data, it is also a minimum of 15 months; and for active archive data, it is approximately 2 – 7+ years minimum.

Please see table 1, which summarizes some of the retention requirements of the regulations and standards that were used in forming the RSA Recommended Best Practices. For more information on the requirements of the regulations and standards and the formation of the RSA Recommended Best Practices, please see the companion white paper, "Log Management Best Practices: The Foundation for Comprehensive Security Information and Event Management," pages 6-9.

Table 1

Industry or Standard	Retention needed to meet audit cycle	Long-term retention of records that may include audit logs
Sarbanes Oxley	1 year	7 years
HIPAA	1 year	6 years
EU Data Protection Directive	1 – 3 years	Not specified
FISMA	1 year	Not specified (refers to NIST standards)
PCI	1 year	Audit trail history must be retained for *at least* 1 year
ISO 17799	Organization-defined	Organization-defined
NIST	1 year	3 years
NERC	1 year	3 years

An information lifecycle management strategy for log data based on combining SIEM with tiered storage
(Reflects RSA Recommended Best Practices for meeting regulatory requirements, audit cycles, and the need for forensic and legal evidence)



* Recommended best practice is one year + one quarter (minimum) on-line to make data accessible for on-going review, real-time analysis and to meet audit cycles

** Recommended best practice is 2-7+ years (minimum) for long-term storage: storing near-line up to 5 years (minimum); possibly move data off-line after 5 years

Infrastructure Requirements

To build an infrastructure for log (event) management that will lay the foundation for comprehensive security information and event management, organizations should consider the following categories of requirements:

- General requirements
- Log generation and capture
- Log retention and storage
- Log analysis
- Log security and protection

Although requirements will vary depending on each organization's particular environment, there are many that are common across organizations. Several requirements for each category are listed below.

I. General Requirements

1. Provides high and consistent performance

Generally, a platform for log management must be able to sustain a high volume of data collection, a high rate of writing to the storage resources without data loss or corruption, a high response rate for retrieving the data, and a high rate of data processing for analysis and reporting. Implementing a high performance infrastructure avoids the prohibitive costs of having to later re-architect for performance.

Performance requirements are determined based on parameters such as:

- The typical and peak volume of log data required to be collected per second (i.e. usually measured in events per second)
- The typical and peak number of users requiring analysis and report generation per hour and per day
- The typical and peak volume of data to be processed for analysis per hour and per day
- The typical and peak volume of reports to be generated per hour and per day
- The typical and peak usage of network bandwidth
- The typical and peak usage of data storage resources
- The security needs for the log data. For example, if log data needs to be encrypted when transmitted between systems, this could require more processing by the systems as well as increased usage of network bandwidth.

2. Enables a distributed deployment

Many enterprises and government agencies today are large, geographically-dispersed and dynamic organizations. The various types and huge numbers of IT systems that are generating log data are spread out across the country or the globe. Different teams such as security or compliance operations often sit at different locations, whether at headquarters or in a particular region. As well, many organizations continue to have more business processes handled at outsourced locations. The number of systems generating logs and the number of stakeholders who need access to the log data are constantly increasing.

These types of organizations are best served by a centrally-managed yet distributed infrastructure that can meet the needs of a geographically-dispersed and constantly evolving organization. With a distributed infrastructure, raw data logs are collected and stored locally on networked storage systems and are then rapidly retrieved and aggregated for analysis and reporting by authorized users. The flexibility of a distributed solution means that the infrastructure can be mapped to any kind of organizational structure. It also means that organizations can quickly adapt to changes as systems or groups of users are added or moved.

Although a distributed infrastructure provides localized collection and storage, stakeholders should be able to get role-based access to the entire enterprise-wide log data set no matter where they are located so that they can perform the analysis and reporting that corresponds to their particular business needs. For example, security operations should have access to enterprise-wide log data for real-time correlation of events and real-time alerts. Compliance teams should also have access to enterprise-wide data in order to generate scheduled and ad hoc reports on overall conformance to policy. In the case of a security breach, the incident management team should be able to perform enterprise-wide queries of the data for a complete forensic analysis.

Another important benefit of a distributed deployment is that it helps to ensure that no data is lost or corrupted as localized collection and storage of the data is faster and more reliable. It also helps in complying with laws that prevent data from being physically moved to another country for processing. While the data is stored locally, specific records can be selectively accessed by authorized users based on content and context.

3. Easily integrates with existing infrastructure

An infrastructure for log management must easily integrate with existing systems to become part of the enterprise's overall IT infrastructure and be manageable within the context of existing operations. The organization should be able to leverage existing systems, for example the SIEM technology should easily integrate with existing storage systems. If new storage systems will be required, the deployment should not negatively impact performance of other systems or create major disruptions to operations.

4. Ensures parallel analysis and storage

An organization should be able to act on the analysis of correlated events while those events are being written to storage. In other words, the platform should be able to support real-time alerts and at the same time be reliably retaining all of the log data as it is collected, so that the data will be available later for audits or forensic analysis.

5. Offers scalability to meet not only current needs but also future needs

A platform for log management must be able to handle average and peak loads. Organizations should plan for surges in the volume of log or event data due to increased activity. The infrastructure also needs to be able to handle increased overall volumes of log or event data due to the addition of more sources of data. Over time, regulations or security requirements may call for more logs to be collected. As well, the organization's IT systems will most certainly grow and more devices will be added. The system should be able to handle increased streams of log data without impacting application performance.

Organizations should also plan an infrastructure that can handle more intensive data processing, analysis, search and retrieval, reporting, and additional storage capacity. The infrastructure will inevitably consume more storage as the business grows and data retention requirements are extended. Look for solutions that offer an efficient growth path as the storage requirements for log data grow.

6. Provides a low total cost of ownership (TCO)

A low total cost of ownership (TCO) is realized through a number of factors. First, it requires minimizing the impact on IT systems not only during deployment but also over time. The time, skill, and effort required by the IT staff to deploy, maintain, upgrade and manage the infrastructure should be minimized; as well as the number of full-time equivalent (FTE) staff required to run the systems. The infrastructure should not require specialized staff such as database or network administrators.

A "small footprint" solution – one that will minimize the impact on IT systems and staff – starts with the deployment of technology that does not require agents to collect then transmit the log or event data to the SIEM server as agents have several inherent disadvantages. They require installation, configuration and on-going maintenance of the agent for every single log source including all network devices, applications and servers. Agents can also be a drain on the host device, especially when multiple agents are present.

Another consideration is that SIEM technology can be software or appliance-based. Appliances are often a much better alternative, offering a smaller footprint solution because installation is simple and maintenance and management are easier.

Since a platform for log management has the potential to use massive amounts of storage resources, to achieve a low TCO, look for a solution that not only avoids generating undue amounts of log data, but also optimizes the use of storage resources.

SIEM technology based on traditional relational database systems (RDBS) can generate extraneous data. In fact, RDBS have the potential to create a data explosion (DE), for example, requiring 12K to 15K of data storage for every 1K of raw log data. This is due to the construction of tables and other overhead. Therefore SIEM technology not based on a relational database will generate less data to store. Also, if SIEM technology uses efficient compression techniques, you can avoid unnecessary purchases of additional storage capacity.

An infrastructure that uses tiered storage also helps to optimize the use of the storage resources. With tiered storage, log data that is infrequently accessed does not take up capacity on primary storage but instead is stored on lower-cost tiers. This reduces the overall cost per MB of storage and puts off the need to acquire additional primary

storage systems. Administrative costs are also reduced as data archived to lower tiers typically requires less management. Freeing up resources on primary storage systems enables read and write requests to be processed much more quickly, which improves overall performance.

7. Supports the retention and retrieval of "evidence-grade" log data

At some point, an organization may need to produce log records to be used as legal evidence or to meet regulatory requests for information. To be used as evidence, the logs should be in the original, unaltered form.

Most laws do not directly address the need to retain the original logs for use as evidence, however there are applicable standards for log management and information security that do. If a court or a regulatory agency expects a company to produce records, the appropriate standard regarding whether the particular logs can be accepted as evidence would likely be a standard such as the National Institute of Standards and Technology (NIST) Special Publications 800-92 "Guide to Computer Security Log Management" or the International Standards Organization "Code of Practice for Information Security Management" (ISO 17799).

Both of these standards indicate that the organization should preserve the original log data for it to be used as evidence.

NIST SP 800-92 states,

> *"The organization's policies and procedures should also address the preservation of original logs. Many organizations send copies of network traffic logs to centralized devices, as well as use tools that analyze and interpret network traffic.* ***In cases where logs may be needed as evidence, organizations may wish to acquire copies of the original log files****, the centralized log files, and interpreted log data, in case there are any questions regarding the fidelity of the copying and interpretation processes. Retaining logs for evidence may involve the use of different forms of storage and different processes, such as additional restrictions on access to the records...".*[1]

ISO 17799 states,

> *"...****To achieve weight of evidence****, the quality and completeness of the controls used to correctly and consistently protect the evidence (i.e. process control evidence) throughout the period that the evidence to be recovered was stored and processed should be demonstrated by a strong evidence trail. In general, such a strong trail can be established under the following conditions:*
>
> *...b) for information on computer media: ...****the original media and the log*** *(if this is not possible, at least one mirror image or copy)* ***should be kept securely and untouched.****"* [2]

When called upon for evidence, organizations must be able to produce log data as part of a discovery process in a reasonable fashion. As well, to provide a digital chain of custody for forensic analysis, you must be able to demonstrate that data has not been altered and that it can, for example, document network usage in an indisputable manner. The infrastructure should capture and store logs in their original format, allowing retrieval of evidence-grade log data for legal, regulatory or forensics purposes.

II. Log Generation and Capture

1. Enables collection of logs from any source and the addition of new sources

Given the variety of systems that most organizations have today, an infrastructure for log management should support collection from all types of security systems, operating systems, and applications. Since it is not always possible to support each and every kind of device or application out-of-the-box, the SIEM technology should also provide an easy way to add support for legacy devices and custom applications.

2. Supports collection of large volumes of data

For SIEM technology, the capacity for processing log data is rated by the volume of events that can be processed in a given time, usually in events per second (EPS). A log management platform should be able to handle typical and peak volumes from all log sources. This includes extreme situations such as widespread malware incidents, vulnerability scanning, and penetration tests that may cause unusually large numbers of log entries to be generated in a short period of time.

[1] Page 10 of NIST SP 800-92 (2006)

[2] Page 105 of ISO 1799 (2005)

3. Performs accurate data collection

For accurate data collection, look for collection methods that do not parse the log messages on input. For example, SIEM technology designed based on a traditional relational database system (RDBS) will parse the log messages in order to put the data into structured columns. With this method, it is possible that data will be incorrectly written to the database or even be lost. This is because log messages from so many devices and systems across the organization will not all follow the exact same format and may not precisely fit the RDBS schema.

If, for example, a firewall has been upgraded, thereby changing the format of the firewall log messages, an RDBS-designed SIEM platform may have difficulties handling the new format. It will possibly mix-up or drop some of the data. If the new log messages contain the information in a different order than the previous version, as data is collected, information previously written to column A may end up in column B, etc. Furthermore, if the new log messages contain additional information that did not exist in the previous version, there will be no corresponding column and the data may be dropped altogether.

Another factor to keep in mind for accurate data collection is how the SIEM technology handles the "pushing" and "pulling" of messages from different kinds of devices. Many UNIX-based devices use the User Datagram Protocol (UDP) to send messages including log data. With UDP push technology, the sender does not find out whether the message has reached its destination and therefore does not retransmit on failure. Thus, the SIEM technology must listen with perfect accuracy and catch every UDP message, even at peak transmission volumes; otherwise the message will be lost permanently.

In contrast, Windows-based devices write event logs to the local disk. With these devices, the SIEM platform must authenticate itself and collect the event log with pull technology using TCP/IP. This approach detects failure and retransmits, but it has higher overhead than does UDP.

To perform accurate data collection, an enterprise SIEM platform must be able to catch and collect every UDP message and be able to handle polling typically thousands of devices three or four times per minute.

III. Log Retention and Storage

1. Supports an ILM strategy whereby data is stored relative to the need for the data using tiered storage

As stated earlier, according to RSA Recommended Best Practices, log data should be stored based on the requirements to access the information given the stage of the data. With tiered storage, data that requires frequent or ready access, such as production data, is stored on-line while data not requiring as frequent or ready access, such as active archive data or back-up data, is stored near-line or off-line respectively.

An infrastructure for log management should include tiered storage that enables the organization to deploy on-line, near-line and perhaps also off-line storage systems offering different levels of accessibility. For example:

1. On-line storage -data is stored on high-performance networked storage systems with access times measured in a few milliseconds providing constant availability to a large number of users

2. Near-line storage – data is stored on a storage subsystem with access times measured in seconds providing for infrequent availability for a small number of dedicated users for very long periods of time

3. Off-line storage – data is stored to disks and tapes that are kept in a data library and cannot be accessed from a computer until mounted

The retention period for production, back-up and archive data will be based on an individual organization's policy. RSA Recommended Best Practices suggest that an appropriate retention period for production data is a minimum of one year plus one quarter. With production data available on-line for a minimum of 15 months, it is readily accessible for real-time monitoring; on-going review and analysis; and in the case of a security breach, for forensic analysis. A 15-month+ on-line retention period also ensures that the complete record of logs is available and readily accessible to meet internal and external audit cycles.

RSA Recommended Best Practices suggest that back-up data be retained near-line for the same period as production data. This ensures that the back-up data would be available in case the production data is compromised or damaged, etc. but it does not need to be as accessible as production data. By deploying a near-line tier, an organization can dramatically improve backup times and efficiency. The sole purpose of backing up the log data is to be able to recover that data when needed. Backing up such data on high-capacity, low-cost disk solutions rather than tape delivers a five-fold improvement in backup and recovery times. It also offers significant reliability improvements to ensure that log data can be recovered quickly and easily when needed.

For active archive data, RSA Recommended Best Practices suggest it be retained approximately 2 – 7+ years (minimum), using near-line storage for more recent records (e.g. up to 5 years), then possibly moving some records to off-line storage (e.g. after 5 years). Although most regulations do not provide definitive time frames for retaining audit logs, they do have general record-keeping requirements of up to many years. Therefore, organizations will typically keep certain logs for several years after an audit cycle. Other requirements for keeping some of the log data longer term include for legal discovery purposes and forensic investigations.

By aligning the storage method with access requirements and retention time frames; and using tiered storage consisting of on-line, near-line and off-line storage; organizations will be able to achieve an ILM strategy for log data that optimizes the use of storage resources.

2. Enables a "cradle to grave" information lifecycle management strategy for log data

The infrastructure should support the management of the entire lifecycle of log data. This means automatically enforcing the organization's retention policies. For example, it should be easy to configure the SIEM platform to store logs for varying lengths of time, including allowing selective retention of logs from different applications for different time periods. The platform should also let administrators easily migrate logs from one storage mechanism to another – such as moving from on-line to near-line storage – and delete logs meeting certain criteria. Retention periods ranging from months to years must be easily managed.

3. Enables fast and fine-grained retrieval of log data regardless of where it is stored (on-line, near-line, off-line)

An infrastructure for log management can grow to contain massive amounts of log data. Being able to provide rapid search and retrieval of the log data makes the system much more useful and the users of the system more productive.

Fast retrieval of stored data logs is facilitated by an infrastructure which:

- Stores logs on-line and near-line: Queries of logs can be significantly slowed by certain storage media, for example loading archived logs from tape instead of directly querying on-line log files.
- Enables centrally-managed data: With a siloed storage approach, data retrieval is slowed because all of the data is not searchable at once; instead it requires searching through log data that is scattered across multiple storage systems.
- Does not use a relational database: Relational databases are too slow to search up to petabytes of data and, because they merge multiple data elements together into rows or tables, they cannot provide fine-grained access to the data elements.
- Implements tiered storage: Removing infrequently accessed data from primary storage systems makes it easier for users to find relevant data quicker.

4. Allows organizations to easily manage log data disposal

Once logs are no longer needed, the platform should provide for easy and reliable disposal. The organization should be able to set up removal of all log entries that precede a certain date and time and be able to permanently delete the information. If information is not properly erased, "deleted" files may be recovered or recreated in the course of litigation, for example.

IV. Log Analysis

1. Provides unified and comprehensive visibility of log information from across the organization

The infrastructure should include a graphical user interface (GUI) that enables review and analysis of all log data aggregated from systems, devices and applications across the network. With a unified view, organizations get a complete picture of their security posture and compliance status. Both real-time and historical events should be presented.

An easy-to-use GUI will ensure the information is readily available and increase the productivity of the analysts. For example, the GUI should assist in quickly identifying potential problems or detecting anomalous events and reviewing all available related data. The level of conformance to policy or compliance to regulations and standards should be displayed in an easy-to-see format. Since every organization will have unique needs, the analysis and visualization tools should be customizable.

2. Detects significant events through correlation

Organizations must be able to readily identify significant events, such as actual incidents or operational problems that necessitate some type of response. Reviewing individual log entries separately is difficult, inefficient, and can create voids. Attacks often involve multiple assets or multiple events. If you watch only individual assets or analyze events in isolation, a single activity may not seem threatening when in fact it is because there are other related activities occurring at the same time. Or another problem is that there are so many events from so many devices being detected, it is difficult to separate the significant events from the insignificant ones.

To filter out the noise and identify the truly significant events, the log management infrastructure must have the capability to perform event correlation – finding patterns of events and looking at relationships between two or more log entries. This is particularly helpful in reducing false positives (whereby the system considers events significant when they are actually not, wasting valuable time and resources as administrators respond to inconsequential events).

The most common form of event correlation is rule-based correlation, which matches multiple log entries from a single source or multiple sources based on logged values, such as time stamps, IP addresses, and event types. Event correlation can also be performed in other ways, such as using statistical methods or visualization tools. If correlation is performed through automated methods, generally the result of successful correlation is a new log entry that brings together the pieces of information into a single place.

3. Generates alerts for all types of attacks and violations

Alerts should be generated to indicate that an identified event or series of events needs further investigation. The alerts must be not only for immediate and obvious attacks, but also for the more stealthy variety taking place over time, anywhere on the network. Often the sophisticated attacker will not hack into a corporate network all at once, but instead mounts what is called a "low and slow" attack by probing for one seldom-used port, then waiting days or weeks probing another, and so on until an opening is found.

The infrastructure should also be able to alert the organization to potential compliance violations as they occur in real-time, such as events that are against the organization's security policy or contrary to regulations and standards. Examples are when a user is attempting an unauthorized access of protected information or when an administrator initiates an unauthorized configuration change. When compliance violations can be detected in real-time, it can greatly reduce the risk of failed audits or penalties.

The alerting system should provide an effective way to prioritize and rank events that represent significant attacks or violations. Otherwise, if the system alerts on everything, it defeats the purpose of alerting since personnel will end up not paying attention.

4. Provides automated baselines

The platform should include the ability to learn the organization's normal operational characteristics and raise alarms if normal parameters are violated. This is especially useful in zero-day attacks, where the signatures necessary to detect the threat do not yet exist.

5. Provides automated and customized reporting

Reporting is performed to summarize significant activity over a particular period of time, record detailed information related to a particular event for forensic analysis, or provide an assessment of compliance to a particular standard or regulation. The platform should be able to handle short-term as well as long-term reporting requirements. For example, if the infrastructure is unable to retain data long-term, it will not be able to meet the requirements for generating reports, say a year after an event, in order to investigate the causes.

To ease the burden of report generation, the platform should provide an ample selection of canned reports, but also offer tools to build custom reports. No two enterprises look alike; therefore security operations and compliance teams should be able to generate reports that are meaningful to their particular organization and actionable by the teams.

6. Facilitates incident management

Most organizations are experiencing a rapid rise in the rate of incidents, which greatly reduces the amount of time available to analyze and respond to each incident. Therefore, an infrastructure for log (event) management should help manage these incidents, including helping to evaluate their significance and formulate a response plan.

Capabilities to look for in SIEM technology that can help with incident management include a "triage" process whereby incidents are sorted, categorized and prioritized. Operations can be greatly simplified when the SIEM platform integrates an incident response workflow; for example, allowing initial assessment and queuing for further handling; providing automated task generation; and tracking the incident information.

Another important area affecting incident management is vulnerability and asset management (VAM). By incorporating information about enterprise assets and known vulnerabilities, a SIEM platform can help reduce the costs associated with incident handling. VAM can help improve the detection of false positives and enable better decision making regarding incidents since analysts get direct insight into the state of an asset (e.g. detected vulnerabilities) and the details of the identified vulnerability.

7. Provides extensive querying and filtering

The infrastructure should ensure that analysts can find the information they need quickly and reliably. They should be able to search all of the log information using user-definable attributes. Another important capability is allowing the filtering of events based on "watchlists," which check for specific user actions, service connections, or known security threats; or attacks coming from known malicious addresses or attacks affecting specific ports, etc.

V. Log Security and Protection

1. Protects data integrity throughout the security information lifecycle

Data has integrity if it can be demonstrated that it has not been altered. The infrastructure should ensure the data is protected from unauthorized alteration and remains accurate from the time it was created through its lifecycle until it is retired.

In general, log data must have integrity: the data must be reliable to ensure that organizations have a true picture of activities within their entire IT environment, and can properly and thoroughly investigate incidents and monitor employee behavior. As well, as mentioned earlier, an organization needs to ensure data integrity in order to meet the requirements of NIST and ISO standards for evidence.

Part of ensuring data integrity in a log management infrastructure is the ability to store logs in a way that reduces the amount of storage space needed for the file without altering the meaning of its contents, (i.e. loss-less compression).

2. Controls Access to Log Data

The infrastructure should prevent users from accessing information they are not authorized to access.

3. Provides for high availability including for log collection, analysis and storage

A platform for log management should be able to ensure there are no interruptions to logging activities so that in the event of a disaster or computer failure, the log collection, analysis and storage services can be restored. Organizations will also need to periodically perform system maintenance or upgrades. The infrastructure should have some way of allowing these without disrupting the logging function.

Conclusion

In today's highly regulated and increasingly risky environment, the need to develop and implement best practices in log (event) management will continue to grow. By providing a complete set of infrastructure requirements, this paper should help organizations realize best practices and formulate a technology strategy that lays the foundation for comprehensive security information and event management.

As organizations are expected to collect and store increasingly massive amounts of log data for longer durations than ever before, an infrastructure for log (event) management must address the need to manage log data over its complete lifecycle. Organizations that develop a competency in log (event) management that incorporates an ILM strategy for log data will have a distinct competitive advantage. These organizations will be able to maximize the value, while minimizing the costs of generating, analyzing, storing, retrieving and ultimately disposing of log data.

Solutions for Implementing Best Practices

Implementing best practices in log management establishes the foundation for comprehensive security information and event management (SIEM). RSA provides end-to-end solutions which enable organizations to build a centrally managed dedicated infrastructure. RSA enVision SIEM platform aggregates logs from across the enterprise and turns this information into actionable intelligence for compliance and security. By combining RSA enVision with networked storage solutions, organizations can manage the entire information lifecycle of logs using a tiered storage approach, whereby logs are stored on different storage resources based on the age of and need for the log data.

The RSA enVision platform works seamlessly with EMC Celerra®, Clariion®, Symmetrix® and Centera® storage for an end-to-end solution in information lifecycle management for log data. This solution enables organizations to manage the huge volumes of log data from creation to deletion in order to meet regulatory compliance, security operations, and business requirements.

For more information on RSA enVision, please go to www.rsa.com. For more information on EMC's storage solutions including EMC Celerra, Clariion, Symmetrix and Centera, please go to www.emc.com.

About RSA

RSA, The Security Division of EMC, is the expert in information-centric security, enabling the protection of information throughout its lifecycle. RSA enables customers to cost-effectively secure critical information assets and online identities wherever they live and at every step of the way, and manage security information and events to ease the burden of compliance.

RSA offers industry-leading solutions in identity assurance & access control, encryption & key management, compliance & security information management and fraud protection. These solutions bring trust to millions of user identities, the transactions that they perform, and the data that is generated. For more information, please visit www.RSA.com and www.EMC.com.

LMBPSIEM WP 0907

RSA Security Inc.
RSA Security Ireland Limited
www.rsa.com

The Security Division of EMC