SANS ANALYST PROGRAM

Sponsored by Alert Logic

Log Management in the Cloud:

A Comparison of In-House versus Cloud-Based Management of Log Data

A SANS Whitepaper – October 2008

Written by: Jerry Shenk

Basic Practices

Questions for the Cloud Provider

Considerations for In-House Log Management

Executive Summary

In the 2008 SANS Log Management Survey, 20 percent of respondents who were satisfied with their log management systems spent more than one week each month on log analysis. Most of those companies were in the Global 2000. The remaining small- and medium-sized businesses (SMBs) and government organizations spent between a half-day to five days per month on log analysis.

The survey also showed that, because of difficulties in setup and integration, most organizations have only achieved partial automation of their log management and reporting processes. These difficulties have organizations, particularly SMBs, wondering if they should turn over log management to an in-cloud provider—one that provides their log management software and log data storage over the Internet. In January, 2008, Stephen Northcutt, president of the SANS Technology Institute, wrote that there are pitfalls with putting log management in-the-cloud. On the plus side, he adds, "you will almost certainly save money. In addition, real experts on log analysis are hard to find..." [1]

Recently, vendors began offering log management in-the-cloud (otherwise known as Software as a Service or SaaS), as a way to simplify log management because the provider can dedicate the material resources and retain the talented, focused personnel to do a better job for less money. This particularly makes sense not only for SMBs without the dedicated manpower, but also for enterprises whose IT resources are stretched trying to manage multiple distributed LANs.

While IT managers agree that log management is difficult, they are leery about handing over their log data to a third party application provider because the data might not be available when they need it, not to mention the sensitive nature of some of the data that shows up in log files.

Before deploying or overhauling log management systems, organizations need to weigh the benefits and drawbacks of each model in context of their business requirements. To simplify the process, this paper presents some questions to consider when vetting those business needs against each (and in many cases, both) of these log management models.

[1] www.sans.edu/resources/leadershiplab/log_logic_interview.php

Basic Practices

When looking at both models of log management (internally or in the cloud), begin with the end in mind by clearly laying out the reasons you want to collect log data. The following are some pre-selection tenets to keep in mind when considering both models of log management:

Identify Your Goals

One of the keys to any successful project deployment is identifying the goals before starting. Log management needs are different for each business unit staking a claim in the process. The IT group may be interested in the value of log data for problem resolution; the security team may be interested in information management or event management tied into an overall SIEM; and the audit and compliance group is most likely interested in tracking what people are doing in regard to sensitive data. Other possible uses for log data include marketing, forensics and HR accounting. As they identify goals, companies would do well to consider the broader advantages of log management and analysis, and look for systems or services that will allow a migration toward a more complete use of log data in the future.

Of importance to all groups is the type of reporting supplied by the service or system. Log management systems often have reporting that is geared toward compliance for PCI, SOX, HIPAA and other similar standards. Apart from required reports, log management can generate reports that are helpful for system maintenance, security management and many other purposes. Whether log management is handled in-house or in the cloud, reporting and correlation features should be easy to use and able to meet current and future business goals.

Locate Resources

Critical to the success of any log management initiative is finding the staff needed to implement, manage and maintain the system. This is particularly difficult for SMBs and government agencies that can't afford top dollar for IT talent. Yet, according to a Gartner paper in May of 2008[2], compliance drivers are pushing organizations with smaller security staffs to acquire log management systems. In these cases, in-cloud services make sense. Larger organizations with dedicated security staffs and advanced log management processes, on the other hand, are more likely to keep log management functions in-house. But even those organizations might use log management services for branches, or as a part of their larger security or network management operations.

[2] "Magic Quadrant for Security Information and Event Management," ID Number G00156945, Mark Nicolette and Kelly Kavanaugh.

Try Before You Buy

The computer industry is fraught with solutions that don't work nearly as well as they purport. So, testing and trial use is critical to determine whether the system or service suits your needs. Put the search interface through its paces to test for functionality and accuracy. Start off with a few devices sending log data, but also set up as many devices as you are allowed to test during the trial period. Some log management systems work very well for a small amount of data; but as the data feed gets larger, the performance goes down quickly—and the systems or services can miss events.

A good way to test the system or service is to send some suspicious data to a device that is being monitored. Then go look for that particular data to make sure it's all there in the logs. One way to do this is to use the Kiwi Syslog Message Generator[3] to send messages to the target, for example by using an option in the program to send a simple text message followed by a number. This makes it simple to see if any of the test messages have been picked up by the log management system or service and reported upon as required.

If there is a security component to the monitoring service (there usually is), try attacking your server and see how the provider responds. The specifics of how you would do this testing will vary with your goals, but logging in as a user and intentionally mistyping the password enough times to lock the account should get a response from the log service or system. I have actively used this testing approach on some appliances that collected security information and never got a response. If you choose to do this kind of testing, start slowly to get an idea of where the response threshold is.

In addition to testing for functionality and security, pay attention to the user interface. In most cases, this will be a Web-based front end. Go through all the options and make sure they work. Also, make sure that responses to the GUI are intuitive. If you have a report that you need regularly, you should be able to get that report reasonably easily, even have it e-mailed to a specified account. Custom reports and specialized reports may be more complicated to receive as a test, but the basic flow of the system should make sense. Finally, make sure that the people who will use the service test the interface before decisions are finalized.

[3] www.kiwisyslog.com/kiwi-sysloggen-overview

Questions for the Cloud Provider

Selecting a log management software service provider is more like cementing a partnership than making a purchase. The service provider will have copies of critical log data—at times they may have the only copies of that data. The table below offers a quick snapshot of what to cover in a Service Level Agreement with a log management cloud service provider. Following that are questions to consider before taking the plunge.

SaaS availability	No more than 2 minutes of downtime a day and no more than 5 minutes per week.
Timeliness of log data showing up in system	Individual logged events must be available to a search from the customer portal within 90 seconds of the event.
Timeliness of log data analysis	Alerts must be delivered to the client within 30 minutes of a critical event.
Regulatory compliance	The SaaS provider must maintain compliance to changing regulations within 30 days of notification of change.
Prompt upgrades to support new attack vectors	New attack vectors should be applied to the processing system within 24 hours of a new attack being identified.
Prompt upgrades to support upgrades to hardware and software	The processing system must be upgraded to support changes and modifications to alerting from supported systems when systems are available for general release.

When considering cloud-based log management applications, organizations should ask the following questions (most of which can also be applied to in-house log management systems):

Is It Safe?

Many IT managers are concerned with the safety of their log data, and rightly so: Log data can be dangerous if it falls into the wrong hands. Attackers can get valuable information from reading the logs. For example, they can see if their attacks work, identify internal hosts, and even identify user names and passwords (which have been known to show up in logs). Log data as common as Web or e-mail traffic often contains confidential information. Having control of logs can be useful to attackers who, in some cases, will try to clean the log data to remove any traces of their activity.

Therefore, it's important to look at the safety of log data—whether it's stored on- or off-site. If the log data is stored locally, it's often kept on each individual computer producing the data. Larger organizations will have log servers that will store the log data in a centralized attached storage device. Those systems are, in an ideal situation, secured and difficult to break into. In the cloud model, this data storage would be handed off to the cloud provider, which relieves the organization of the hardware, security and HR burdens involved with keeping storage in-house.

However, as they lose control of that data, organizations must rely on the cloud service to handle their data securely. The issue of whether a service organization is competent is difficult to determine, and is ultimately based on reputation. Cloud providers must create a trust model as they manage collected log data securely and separately in a multi-tenant environment. This creates the need for additional layers of security to separate multiple tenants from one another on a shared server, while also protecting the data stores from attackers.

Firewalls, encryption and data loss prevention are all areas of security that apply to sensitive log data stored in the cloud—a cloud that's increasingly virtualized. Virtualization, in itself, is not necessarily a negative, as long as proper security procedures are followed within the virtual cloud. The same characteristics of virtualization that make it a concern as a hacking agent also provide a hiding technology that has the potential to make user accounts harder for attackers to access. Already security vendors are developing virtual technologies so that their anti-malware products can't be detected and overruled by today's kernel- and boot-level rootkits.

How Is It Transported?

Ask the cloud provider for specifics about how the data gets transmitted from your systems to their operations center. Is the data encrypted in transit? What type and strength of encryption is used? Is the encryption proprietary? Be wary of providers that claim their encryption information is confidential or proprietary; instead, look for providers that use proven technologies such as SSL or AES. There are numerous examples of companies that have invested vast amounts of money in creating their own encryption technologies only to find out after release that they missed a critical component.

How Are Keys Stored?

It would be easier for a log management vendor to use the same encryption secrets for all their clients; however, that means that if an attacker discovers a key to one client, that attacker can access the accounts of all clients. A different key for each customer account would not only offer better protection against customers accessing one another's accounts, but also against an attacker cracking a password and getting the keys to the entire kingdom. Logical separation of key storage is also important for the same reasons.

How Often Is The Data Transmitted?

Most log management systems send data in batch mode. The collection appliance typically waits for either a specified time or amount of data before transmission. In general, a quicker frequency is better because the data is getting processed faster. More frequent transmission minimizes traffic bursts and gives an attacker less time to interrupt or block the transmission of alerts, a technique attackers use in an attempt to avoid detection.

What Is The Level Of Compression and Bandwidth Utilization?

Bandwidth utilization is a question that you'll want to keep an eye on as you test your log management service. It is common to get 90 percent compression or better on ASCII (plain text) logs, while binary log compression ratios may be less. If your Internet connection is currently heavily utilized, the log traffic may impede other traffic, and you'll want to plan for this issue ahead of time. One way to monitor the bandwidth is to capture traffic statistics using Net Flows. If you aren't monitoring your overall Internet traffic utilization, it's best to get a handle on that prior to implementing a log management service and use this number as a baseline.

What Backup and Redundancy Is Included?

If a cloud provider claims to be set up to handle this type of data correctly, verify that it is, in fact, doing a better job than you would. The provider should have data stored at multiple locations and secure backups for redundancy. Check, too, with the company that is actually doing storage. In the cloud model, storage could be handed off to another vendor.

Ask questions about how stored data is encrypted, how it is transferred, where the tapes or other media are stored, and if there is a process for tracking tapes. Find out how long backup data is retained, how it's destroyed, and what happens to the data if the service is terminated. It will probably be impossible to verify most of this, but the cloud provider should be able to answer questions and provide benchmarks, customer references, service agreements and other documentation.

What Are The Responding Options?

Log management systems typically provide reports that are already built into the system. These built-in reports typically cover things like regulatory compliance and common performance and security metrics. Verify that the reports your organization needs are included as prebuilt reports, or that they're easy enough to customize.

Often, reporting is not as straightforward as people would like it to be. Sometimes, the logging application won't provide the required information directly, but it may be available indirectly. For example, a security manager may want to identify invalid session IDs on a Web site because a high frequency of invalid session IDs may point to an attacker trying to guess a session ID and clone or hijack the session. If the log manager doesn't report that information directly, it may be possible to get similar information by tracking the number of connections built from any given IP address.

How Much Of The Data Is Actively Searchable?

In some cases, the most recent data will be more quickly accessible for searching than data that has been removed from an active state. Moving data out of an active part of the database can make databases faster, so some data may be moved into an area that provides slower access. Ask if there are any special requirements to access archived data or whether the only issue is a performance penalty—and request a demonstration.

How Much Of The Data Is Stored?

If data is moved out of primary storage, is the full log data retained or will recovery of data be limited to searchable data fields and metadata? If some detail is eliminated, determine whether this will cause problems with regulatory compliance, forensics processes and other log management and reporting needs required across your organization.

If there are processes that automatically eliminate some data, can those processes be suspended for special circumstances, such as litigation requiring the preservation of data? How long does it take to make such changes?

What Log Data Will Be Accepted?

What specific devices, operating systems and applications are supported? Several operating systems and hundreds of widely used appliances and devices are critical to today's diverse organizational IT infrastructures. The number of applications a log manager may be called upon to understand is staggering. Prioritize on all your critical devices and applications. How are they supported by the service provider, and how thorough is that support?

How Are Its Instructions For Setting Up Devices?

Log management can become more complicated as the number of log-producing devices increases. An in-the-cloud log management provider should have guidelines for setting up devices, operating systems and applications that need to be monitored.

Often, a company will need to deviate from the normal setup procedure, based on the peculiarities of its business that complicate the log data life cycle. As a result, setup instructions should be termed as guidelines, not hard and fast rules. Rules often must be massaged to work with the varying operating systems and applications (including their versions) that an organization needs coverage for.

How Are Alerts Determined?

If the cloud provider is offering to send alerts for events of interest, find out how they determine what is of interest and compare that to what is of interest to your organization. Are they looking solely at security events or do they include more routine support and maintenance events? If the events of interest include both types of events, how do they distinguish between the two? How much involvement does the log management client have in setting up what alerts are of interest to them? If a drive runs out of space, for example, that can often be just as big a problem as an attacker compromising a system.

Ask, also, if they can correlate related events to give the analysis situational awareness. For example, an administrator logging into a domain controller at 10 a.m. and creating a user is quite different from the DNS process starting a command shell and creating a user in the middle of the night. In both cases, a user is being created. In the first instance, the process seems normal; but in the second instance this combination of events could be associated with the RPC DNS exploit as demonstrated in an April, 2007, SANS Webcast[4]. Cloud (and in-house systems) should, therefore, include situational awareness to understand when creating a user is a normal event and when, as in the second example, it is not normal.

In addition to automated monitoring and alerts, it would be ideal if cloud providers could offer human review of logs as an add-on fee for service. Human review is required under some regulations, and is a good basic best practice for organizations to follow because automated systems don't catch everything.

How Quickly Does Processing Occur?

Timing is an important issue with log management that the cloud model is well-suited to address. One typical problem with in-house log management is that events are often found after a problem is noticed. It is, of course, best to detect log events leading up to a critical event in order to avoid the critical event.

The question about processing speed encompasses a number of different issues: Once an event has been logged at the local device, how long does it take for that event to show up in the system? If that event should trigger an alert, how long will it be before the alert is relayed to the client IT department? Is there an option for the vendor to do more than just shoot out an e-mail message? How is the escalation process handled?

[4] April 24, 2007 Webcast - www.sans.org/webcasts/show.php?webcastid=90861

How Often Are The Alerts Updated?

Operating systems and network devices are constantly coming under new and different attacks requiring new responses. The errors from these devices also change with some upgrades, so it is important for the Log Management provider to conduct regular and timely updates to its system, and respond reasonably when errors occur.

How Are System Upgrades Handled?

In the cloud, upgrades to the log management systems are handled by the provider, thereby relieving the organization from having to maintain these systems in-house. There is a risk, however, that the upgrades may cause outages in coverage by accidentally introducing new compatibility or protocol problems. It would be a good to ask the cloud provider about how upgrades are handled and how clients are protected during the upgrades. By the same token, how would updates to any internal system log-generating devices affect the cloud provider's coverage?

Considerations for In-House Log Management

Many of the same questions that apply to companies offering log management service in the cloud also apply to internally-managed log management systems. The 2008 SANS Annual Log Management survey indicates it is still incredibly difficult to automate log management systems to the degree organizations need. A recent article by Patrick Mueller in *Information Week*[5] refers to log management as a "monster." Just because it's difficult doesn't mean log management needs to be outsourced. When weighing in-house log management, consider the following factors:

Could A Personal Change Ruin Your Log Management Process?

Log management is often the pet project of one person while the rest of the IT staff tries not to get involved. If that person leaves the company, it can be difficult for someone else to pick up the task and keep it going. This is a problem with many IT initiatives.

Will Your Staff Monitor The Logs Regularly And Maintain Updates?

Log management services have requirements built into their contracts for response time and full-time monitoring. Can your staff live up to those same expectations? One of the issues for log management companies is keeping up with updates to applications, operating systems and regulatory issues. Is your staff able to keep up with the changes? As an example, how did your staff do when Windows Server 2008 changed all its event IDs[6]? At the time, most administrators used a collection of scripts; however, all those scripts, which were working, suddenly became broken. Bloggers lashed out about it. For a log administrator who finally has everything working, that sort of a situation can be a demoralizing surprise.

Maintaining updates and monitoring logs is complicated by the fact that most companies support a diversity of logging devices. To properly support local log management, an IT group will need to work with different vendors who use different types of log data. At times, it may be necessary to bring in consultants to assist with tracking down specific issues. Organizations need to consider the associated costs and frustrations of working with multiple vendors and integrators along with the costs of the initial deployment and ongoing internal staffing requirements.

[5] www.informationweek.com/story/showArticle.jhtml?articleID=208400730
[6] www.ultimatewindowssecurity.com/wiki/WindowsServer2008VistaSecurityLog.ashx

Roll Your Own Or Buy An Appliance?

A big debate in the log management arena is how to deploy log management tools. According to the SANS Log Management survey, the majority of organizations (38 percent) are building home grown solutions through the use of syslog servers, custom scripts and applications. The remaining respondents used a combination of commercial software and appliance-based tools or checked "other." In either case, organizations are not happy with their level of automated correlation or system coverage, according to the survey. Coverage, automation, correlation and access must all be addressed, maintained and improved upon as needs dictate, regardless of which option is chosen.

Summary

Log data can provide valuable security and operations insight into enterprises and SMB IT operations. Many companies with limited IT staffing will find that outsourcing log management can bring them more value from their log data than they could attain on their own—without all the expenditures in hardware, staffing and product management. If in-cloud providers can deliver prompt, secure, reliable service, cloud-based log management could be a growth sector over the next few years, particularly for the SMB market.

One of the big questions in making a decision between internal and in-cloud log management is how much time can be allocated to monitoring and upgrading the system internally to meet the business needs of the organization. If an organization's primary goal is regulatory compliance or to minimize IT staff requirements, then outsourcing log management to cloud application providers will probably be suitable.

Organizations that decide to outsource their log management should be careful to select flexible services that allow for expanded correlation and use of the log data for organizational benefit. This is also true of internally-developed log management systems, which today are experiencing interoperability issues that make data normalization and correlation difficult for organizations of all sizes.

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: