

Log Management for Marshall University: Issues and Recommendations

1 Introduction

Most components of an information technology infrastructure are capable of producing logs chronicling their activity over time. These logs often contain very detailed information about the activities of applications and the layers of software and hardware that support those applications.

Examples include:

- Application logs can identify what transactions have been performed, at what time, and for whom. Those logs may also describe the hardware and operating system resources that were used to execute that transaction.
- Network devices, such as firewalls, intrusion detection/prevention systems, routers, and switches are generally capable of logging information. These logs have value of their own to network administrators, but they also may be used to enhance the information in application and other logs.

With proper management, these logs can be of great benefit in a variety of scenarios, to enhance security, system performance and resource management, and regulatory compliance. In particular, a *log management infrastructure* can capture information and aid analysis about the following:

- *Access*. Who is using services.
- *Change Monitoring*. How and when services are modified.
- *Cost Allocation*. When applicable, who should pay how much for services.
- *Malfunction*. When services fail.
- *Resource Utilization*. How much capacity is used by services.
- *Security Events*. What activity occurred during an incident, and when.
- *User Activity*. What people are doing with services.

The following are example scenarios illustrating how the information in logs can be critical to resolving a security or operational issue:

- Internet access is very slow for 10-20 minute periods at random times throughout the day. Router logs identify a high rate of transmission errors on the campus's Internet connection at those times. The network administrator calls the campus's ISP to repair the connection.
- Internet access is very slow, but everything seems to be working correctly. Firewall and router logs determine that a particular PC in the dorms is under attack from over 500 addresses located around the world, most likely a botnet. This allows the network administrator to have that PC's Internet traffic blocked, relieving the campus's network-wide congestion.
- A student has been dropped from all of her classes and accuses the Registrar and threatens to sue. An application log shows that her user ID had been used to drop the classes two hours after she had dropped her boyfriend, and that the session originated from her boyfriend's dorm room. She remembers that she had given her password to her ex.
- A server containing sensitive information (Social Security numbers) was compromised by hackers who installed an FTP service and pornographic images. To determine who, how and extent of the compromise the logs from the server operating system, the FTP software, firewall and IDS system are all analyzed as part of the investigation.

The University must manage these logs in a manner that facilitates these benefits while protecting the privacy and integrity of the information contained in these logs. This document recommends appropriate practices for log management within a campus's overall IT infrastructure. It should, however, be noted that While Marshall University Computing Services makes a best effort to retain logs for the recommended periods. Networking/service interruptions, hardware failure, software failure, and limited resources may impact UCS ability to obtain accurate logs and retain them for the recommended periods.

2 Candidate Sources of Logged Information

There are many sources of log information:

- Application logs have the potential to identify what transactions have been performed, at what time, by whom, and on what object. Those logs may also describe the client and server hardware and operating system resources that were used to execute that transaction.
- System logs for operating systems and services, such as web, database, authentication, print, *etc.* provide detailed information about their activity and are an integral part of system administration. When related to application logs, they provide an additional layer of detail that is not observable from the application itself. Service logs can also aid in intrusion analysis, when an intrusion bypasses the application itself.
- Network devices, such as firewalls, intrusion detection/prevention systems, routers, and switches are generally capable of logging information. As before, these logs add purpose of their own to network administrators, but they also can enhance the information in operating system, service, and application logs.
- Change management logs that document changes in the IT or business environment provide context for the automatically-generated logs described above.
- Other sources, such as physical access or surveillance logs can provide context when investigating security incidents.

These logs have the potential of being very large, depending on the volume of activity and the amount of information in each log entry. The cost of storage and processing should be considered when determining which sources should be incorporated into the log management infrastructure, as well as the potential benefit of having that information at some time in the future.

3 Recommended Log Management Practices

3.1 Log Generation

This section describes information that might be included in various types of logs. It should be noted, however, that the information in logs often cannot be controlled by application, system, or network administrators, so the items listed here, while often highly desirable, should not be viewed as absolute requirements.

3.1.1 Application Logs

Applications should log their activity in a manner that correlates well with the business processes the applications support, particularly any operations that modify permissions or access rights. These logs should include:

- The business operation that was requested
- Whether the request was accepted or denied
- The time and date the operation was performed (Start and end times may be appropriate for long operations.)
- Who initiated the operation

- System and network resources used
- Any information needed for business process controls
- Client hardware and software characteristics

It should be noted that the "application" may actually be a more generic service, such as a web, file, or print server, or even a PBX. In this case, it may be difficult to relate the more generic logs to business processes. When this is the case, appropriate documentation may need to be maintained describing the relationship between the logs and the supported business processes.

3.1.2 System Logs

Many components of the IT infrastructure generate logs. Examples of these components include:

- Operating Systems
- Web servers
- Database servers
- Print servers
- File servers
- Authentication servers
- DHCP servers
- DNS servers
- Electronic mail server logs

In general, all of these logs have potential value and should be maintained. These logs should include the following types of information:

- The server operation that was requested
- Whether the request was accepted or denied
- The time and date the operation was performed (Start and end times, or duration, may be appropriate for long operations.)
- Who and/or what system initiated the operation
- System and network resources used

Host-based firewalls also generate valuable log information. This is described in the "Network Logs" section of this document.

It should be noted that client workstations also generate system logs that are of interest, particularly for local authentication, malware detection, and host-based firewalls

3.1.3 Network Logs

Many components of the network infrastructure generate logs. Examples of these components include:

- Routers
- Switches
- Wireless access points
- Network-based firewalls
- Host-based firewalls
- Intrusion detection and prevention systems
- Telephone Switches

These logs typically describe *flows* of information through the network, but not the individual packets contained in that flow. (A *flow* is the traffic that corresponds to a logical connection between two processes in the network. Examples of flows include a connection to a web server, a remote login session, or a Domain Name System lookup.) Information logged for a flow should include:

- Network (IP) addresses or telephone numbers of the end points
- Service identifiers (port numbers) for each of the end points
- Whether the flow was accepted or denied
- Date, time, and duration of the flow
- Number of packets and bytes used by the flow

Other components of the network infrastructure, such as DHCP and DNS servers, provide valuable information about network configuration elements, such as IP addresses, that change over time. The logging requirements for these servers are covered in the "System Logs" section of this document.

3.1.4 Time Synchronization

One of the important functions of a log management infrastructure is to relate records from various sources by time. Because of this, it is important that all components of the IT infrastructure have synchronized clocks. Use of a time service, such as NTP, is highly recommended.

3.2 Use of Log Information

3.2.1 Baseline Behavior

It is essential that a baseline of activity within the IT infrastructure be established and tracked as it changes over time.

- For system and network administrators, this should include the volume of activity for major applications and systems, as well as traffic volume over the network, and should be presented over a common time scale.
- It may also be desirable to present application activity to business managers in a manner that enables them to correlate the information with business volume.

Understanding baseline behavior allows for the detection of anomalous behavior, which could indicate a security incident or a change in normal usage patterns. Procedures should be in place to ensure that this information is reviewed on a regular and timely basis.

3.2.2 Investigation

When an incident occurs, various *ad hoc* questions will need to be answered. These incidents may be security related, or they may be due to a malfunction, a change in the IT infrastructure, or a change in usage patterns. Whatever the cause of the incident, though, it will be necessary to retrieve and report log records based on a variety of selection criteria, such as:

- Source(s) of the log records
- Time
- Network address
- Application or service
- User

When matching records from multiple sources, time and network address will be the most valuable for matching records. Application, service, and user may also be desired for matching, but it is likely that they will need to be associated with network address and time in order to accomplish this.

4 Appropriate Use of Log Information

Logs often contain information that, if misused, could represent an invasion of the privacy of members of the MU community. While it is necessary for the University to perform regular

collection and monitoring of these logs, this activity should be consistent with the provisions described in MU's Information Technology Acceptable Use Policy.

5 Log Record Life-Cycle Management

When logs document or contain valuable information related to activities of the University's information resources or the people who manage those resources, they are University *Administrative Records*, subject to the requirements of the University Records Management Program to ensure that they are appropriately managed and preserved, and can be retrieved as needed. See the University Business Records Retention Policy, particularly:

- Records Retention and Disposition, and
- Legal Requirements on Privacy of and Access to Information.

The following are specific issues that should be considered with respect to such log records.

5.1 Retention

In order to facilitate investigation as well as to protect privacy, the retention of log records should be well-defined to provide an appropriate balance among the following:

- confidentiality of specific individuals' activities,
- the need to support investigations, and
- the cost of retaining the records

The records required to support investigation often contain information about specific individuals' activities and must, therefore, be protected adequately against unauthorized disclosure. It is also the case that records required for long-term analysis often contain information about specific individuals' activities, but that specific information is not needed. When the retention period required for analysis is significantly longer than that for investigation, new records can be created by aggregating or redacting the original records to reduce the cost of managing the records and protecting them against unauthorized disclosure.

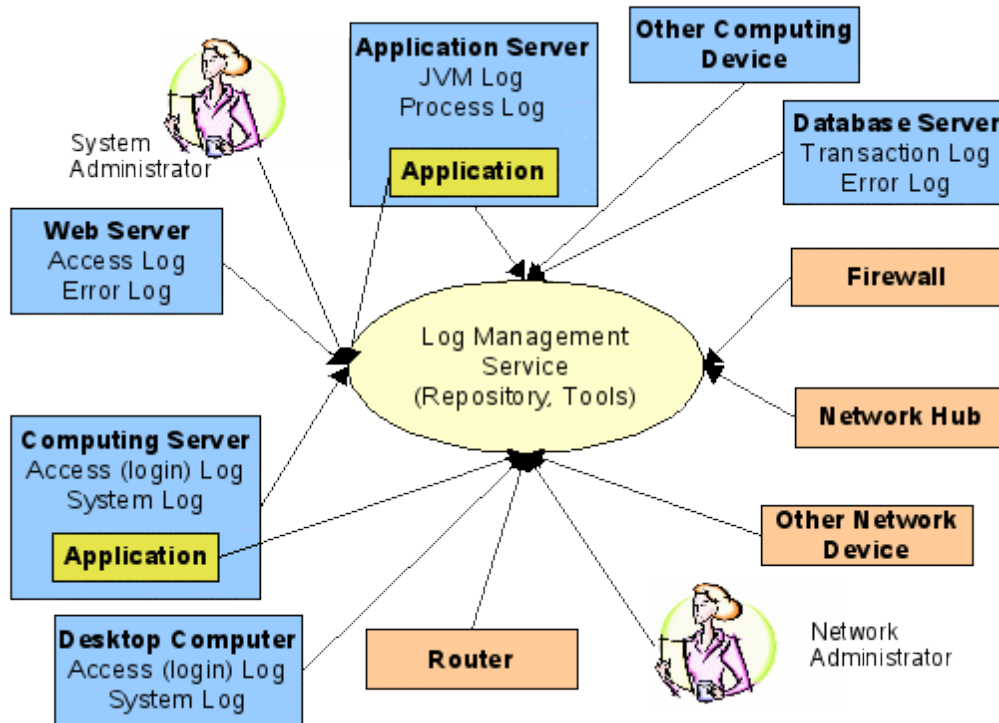
Care should be taken not to retain log records that are not needed. The cost of long-term retention can be significant, and could expose the University to high costs of retrieving and reviewing the otherwise unneeded records in the event of litigation.

5.2 Log Management Infrastructure

As a best practice, a log management infrastructure should be established to provide common management of log records. This infrastructure will:

- move log records into the infrastructure,
- provide secure storage for the records,
- implement record retention policies,
- facilitate access to log records,
- provide analysis tools that enable correlations among records from multiple sources, and
- protect the chain of evidence for the possibility that log records are used in legal proceedings.

The following diagram illustrates such an infrastructure.



In order to facilitate the creation of campus-based log management infrastructures, it is recommended that system-wide groups be established to address the following issues:

- Technology solutions that can be used to build log management infrastructures.
- Typical retention periods for common examples of logged information.

6 Summary of Recommendations

Section	Recommendations
Introduction	The University should manage logs in a manner that facilitates the benefits described in this document, while protecting the privacy and integrity of the information contained in these logs.
Candidate Sources of Logged Information	The cost of storage, processing, management, etc., as well as the benefit benefit, should be considered when selecting sources for incorporation into the log management infrastructure.
Application Logs	Applications should log their activity in a manner that correlates well with the business processes the applications support, particularly any operations that modify permissions or access rights. These logs should include, at a minimum: The business operation that was requested Whether the request was accepted or denied The time and date the operation was performed (Start and end times may be appropriate for long operations.) Who initiated the operation System and network resources used Any information needed for business process controls Client hardware and software characteristics

System Logs	<p>System logs should include the following types of information:</p> <ul style="list-style-type: none"> The server operation that was requested Whether the request was accepted or denied The time and date the operation was performed (Start and end times, or duration, may be appropriate for long operations.) Who and/or what system initiated the operation System and network resources used
Network Logs	<p>Information logged for a network flow should include:</p> <ul style="list-style-type: none"> Network (IP) addresses of the end points Service identifiers (port numbers) for each of the end points Whether the flow was accepted or denied Date, time, and duration of the flow Number of packets and bytes used by the flow
Time Synchronization	<p>One of the important functions of a log management infrastructure is to relate records from various sources by time. Because of this, it is important that all components of the IT infrastructure have synchronized clocks. Use of a time service, such as NTP, is highly recommended.</p>
Baseline Behavior	<p>The baseline of activity within the IT infrastructure should be established and tracked as it changes over time.</p> <p>For system and network administrators, this should include the volume of activity for major applications and systems, as well as traffic volume over the network, and should be presented over a common time scale.</p> <p>It may also be desirable to present application activity to business managers in a manner that enables them to correlate the information with business volume.</p> <p>Procedures should be in place to ensure that this information is reviewed on a regular and timely basis.</p>
Investigation	<p>When conducting an investigation, it will be necessary to retrieve and report log records based on a variety of selection criteria. Preparations should be made to perform ad hoc queries based on criteria, such as the following:</p> <ul style="list-style-type: none"> Source(s) of the log records Time Network address Application or service User <p>When matching records from multiple sources, time and network address will be the most valuable for matching records. Application, service, and user may also be desired for matching, but it is likely that they will need to be associated with network address and time in order to accomplish this.</p>
Appropriate Use of Log Information	<p>While it is necessary for the University to perform regular collection and monitoring of these logs, this activity should be consistent with the provisions described in MU's Electronic Communication Policy.</p>
Retention	<p>In order to facilitate investigation as well as to protect privacy, the retention of</p>

	log records should be well-defined to provide an appropriate balance among the following confidentiality of specific individuals' activities, the need to support investigations, and the cost of retaining the records
Log Management Infrastructure	Each campus should establish a log management infrastructure to do the following: move log records into the infrastructure, provide secure storage for the records, implement record retention policies, provide analysis tools that enable correlations among records from multiple sources, and protect the chain of evidence for the possibility that log records are used in legal proceedings. System-wide groups should be established to address the following issues: Technology solutions that can be used to build log management infrastructures. Typical retention periods for common examples of logged information.

Log Class	Type		Retention period
Application			
	ERP (Banner)		90 days
		?	90 days
	Portal		90 days
	Web		90 days
	...		90 days
System			
	Operating Systems		90 days
	Web Servers		90 days
	Database Servers		90 days
	Print Servers		90 days

	File Servers		90 days
	Authentication Servers		90 days
	DHCP Servers		90 days
	DNS Servers		90 days
	Email Servers		90 days
	...		90 days
Network			
	Routers		90 days
	Switches		90 days
	Wireless APs		90 days
	Firewalls		90 days
		PIX	90 days
		chassis	90 days
	IDS/IPS		90 days
		Cisco Clean Access	90 days
	Telephony		90 days
	...		90 days
Workstation			?

7 Acknowledgments

This document is in large part the result of work done by a subgroup of the University of California Information Technology Policy and Security Officers. The work group members were:

Jacqueline Craig, UC Office of the President

Jon Good, UC Office of the President

Karl Heins, UC Office of the President

Binh Nguyen, UC San Francisco Medical Center

Carl Tianen, UC San Francisco

David Walker, UC Office of the President

May 1, 2006

Inserted from <<http://www.ucop.edu/irc/itsec/uc/LogManagementGuidelines-2006-05-01.html>>