context

Effective Log Management

www.contextis.com

CPNI®
Centre for the Protection
of National Infrastructure

About

About Context Information Security

context

Context has a client base including some of the world's most high profile blue chip companies and government organisations. Our strong track record is based above all on the technical expertise, professionalism, independence and the integrity of our consultants. Our comprehensive portfolio of technical services sets the standard for the security industry.

Context has a dedicated incident response and investigation team, which works with clients to detect, investigate, understand, respond to, and protect against nefarious activity on their networks. This team does not focus on generic malware, but offers services designed to counter the most sophisticated attacks targeted against our clients. Our team comprises experts who can provide substantial experience and expertise on a business as well as a technical level.

Context is proud to be part of the Cyber Incident Response scheme run by CPNI and CESG, the Information Assurance arm of GCHQ.

About This Booklet

Context would like to acknowledge the help and support of CPNI in researching this topic and producing the accompanying products.

Introduction

Log files are historical records of the running state of hardware and software, storing information on how they are used, errors that occur and application-specific events which detail how users interact with them. Routine review of this information can provide system administrators and computer security teams with insight into how effectively the business is operating and where configuration errors may be causing issues on the network so that they can be remediated before they have wider impact.

Log records are also an immensely valuable source of information for computer security purposes, but their value as part of a corporate intrusion detection and incident response process is largely misunderstood by many organisations; logs are either not collected at all or are collected without consideration for how they might be used should an incident occur. The Effective Log Management project has been commissioned by the UK Centre for the Protection of National Infrastructure (CPNI) to demonstrate the value of this data to the reader and discuss how it can be used to support an efficient response to a network intrusion. This booklet accompanies the full whitepaper "Effective Log Management", available from **www.cpni.gov.uk**. The advice within these documents is intended to be used in conjunction with the 20 Critical Security Controls for Effective Cyber Defence coordinated by the Council on CyberSecurity.

Only 6% of network breaches discovered by log review

While 86% of victims held evidence of the breach in log files*

Developing a log management strategy to enhance an organisation's computer security posture is not easy and there are a number of hurdles to overcome. The number of devices that generate logs on a network can be overwhelming and storage strategies can be both expensive and complex to implement. What logs should be collected? For how long should they be stored? How should they be stored to best ensure the security and integrity of the data? How can log files be used proactively to look for the malicious "needles" in the "haystack" of data available?

The lesson for organisations is simple: logs are an evidence source of potentially vital importance, but effort is required to exploit them for maximum value. They are a dataset which requires little effort to start collecting at some level, but are greatly helpful during the investigation of security breaches. They can be used by forward facing organisations to identify, track and mitigate attacks before any damage is caused. However, in all cases a clear log management strategy is essential to ensure the data is of sufficient quality and is organised appropriately.

* source: http://www.verizonenterprise.com/resources/reports/rp_2010-data-breach-report_en_xg.pdf

Challenges

When defining a log management strategy within an organisation, a balance will likely have to be struck between breadth and depth of data collection and how it will be stored and analysed. While the ideal policy may be to fully log all possible data sources and retain this information for extended periods whilst simultaneously analysing them for anomalies, this is rarely practical for many organisations. Therefore, it is critical that the inclusion of each data source is blended into an overall incident response process to create a layered approach that ensures that the best possible view of an intrusion can be determined as quickly as possible

Common challenges associated with effective log creation, analysis and management include the number of devices capable of generating log data and the associated resources required for retention, the security of log data storage and the effective analysis of data once it has been captured. An additional challenge is the involvement of third-party IT service providers and their ability to supply responders with the required data in a timely fashion.

Comprehensive logging over the longest timescale possible will allow incident responders to generate a detailed picture of an intruder's activities. Many compromises are only discovered weeks, months or in extreme cases, years after the initial breach.

Top three primary challenges to attack detection using logged events:

1. Organisations not collecting the appropriate operational and security-related data to enable analysis
2. A lack of system and vulnerability awareness for network assets
3. A lack of relevant event context to observe "normal behaviour" and detect anomalies

(Source: https://www.sans.org/reading-room/analysts-program/security-analystics-survey-2013)

Advantages

The Role of Log Management in Incident Response

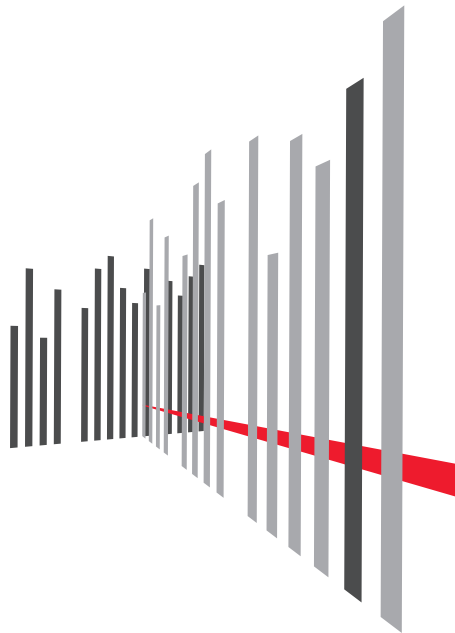
Correlation of log events across a range of devices is a critical part of any incident response activity as it assists an organisation in assessing the extent and impact of a network compromise, as well as informing what steps may be necessary for mitigation. Network security appliances may provide elements of the picture during an attack, but may not have full visibility over a network or possess the correlation capabilities to fully describe the attacker's activities.

For example, a security product may identify when an intruder performs a remote attack against an organisation's web application, but it may not have the alert rules to identify the attacker's propagation through a network from the initial point of compromise. By correlating the original alert event with log files from the web, database and authentication servers as well as events generated on the hosts, greater visibility into the extent of the compromise may be established.

The Role of Log Management in Intrusion Detection

While log files are typically used to respond to incidents by determining the 'how, when and what' of an intruder's activities, they can also be used proactively to monitor the security posture of a network. By analysing the data generated by devices on a regular basis, security teams may be able to detect and respond to the initial stages of a security incident as they occur.

Intrusion detection requires greater up-front resource commitment both in staffing and time allocation, but proactive analysis can detect events that may not be highlighted by automated systems and can be used to limit the wider impact and cost of a security incident.



Log Management Preparation

The generation, storage and analysis of log data must be defined by a process that first identifies the assets that need protection, establishes data sources that can contribute to defending these assets and then validates their inclusion in the log management strategy. Validation of log sources should be iterative and be informed by the process of analysis; if the captured data is of limited value or stores too many records for the projected retention period, the inclusion of the data source as it currently exists and length of time the data is stored for should be re-evaluated and incorporated back into the overall strategy.

Log Sources

While log data can be generated by large numbers of devices, they generally relate to one of four source classes:

- Services that provide functionality to users
- Infrastructure supporting the network
- Host devices
- Remote connection services

The focus of this booklet is to identify appropriate sources of log file data within an organisation in order to facilitate intrusion detection and incident response processes. An increase in the availability and popularity of 'cloud' and managed services means that logging for these data sources may not be readily controlled by local network administration. The following recommendations will still be relevant in situations where a third-party supplier provides a service, although the availability and accessibility of particular records may vary. Therefore, it is important that the availability of log data from these services is discussed and agreed with the provider in advance of potential security incidents.

For all log sources, both "failures" and "successes" should be logged wherever possible, as some attacker activities may not be considered anomalous and therefore would not be stored as exceptions. Without logging successful actions, it will be extremely difficult to compile an accurate picture of how an attack was executed and what follow-on activities were conducted.

Without logging successful actions, it will be extremely difficult to compile an accurate picture of how an attack was executed and what follow-on activities were conducted

When establishing a log management policy, it is important to consider what data is appropriate to store at a minimum to enable intrusion detection and response procedures. While storing all data for every possible source may be the ideal for full response purposes, this might not be practical in every situation. The following pages provide a product-agnostic view of the primary log sources on a typical network, describing how they might be used during analysis to contribute to the overall picture of a security incident.

Servers

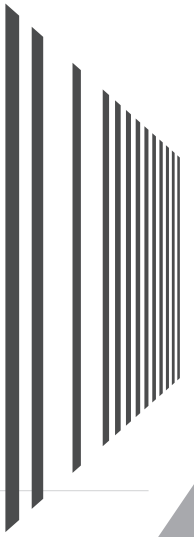
Proxy Server

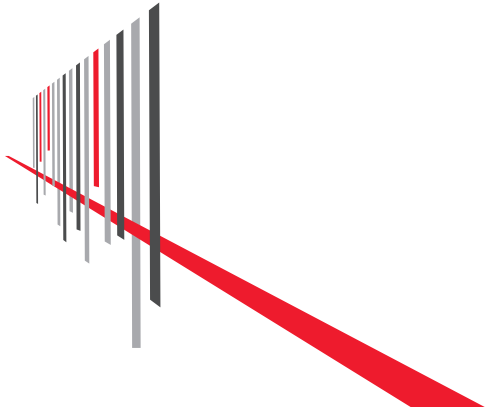
Proxy servers are intermediate devices through which remote hosts such as web servers are accessed from within a network. If employed on a network, web proxies can provide a source of log data for outbound network traffic as they store information on requests to web servers including the Uniform Resource Locator (URL) requested and the status of the response. This information can provide an indication of outbound requests sent from malware to obtain tasking and to send data back to an attacker, along with the IP address of the affected host on the network.

Mail Server

A common attack vector for host compromise is spear phishing or "content delivery" attacks, where an attacker sends emails to a target containing malicious software as an attachment. Logging of metadata and content pertaining to inbound email traffic can help establish which users have received emails with malicious content, when the attacks occurred and other information that may be useful for ongoing detection purposes, such as email subject lines that may be reused or could provide insight into the methodologies used to entice users into opening the attachments.

In some cases, malicious software may have the capability to use Simple Mail Transfer Protocol (SMTP) as a command and control or data exfiltration channel. When investigating incidents involving this type of malware, logs of outbound email sessions can also be useful to determine the frequency and type of data sent from the network.





Servers

Web Server

Web servers represent a potential attack vector for an attacker either via the exploitation of vulnerabilities in the applications they are serving or through the web server software itself. Servers that are connected directly to an external network such as the internet can provide an initial foothold into a network, whereas internal web servers may be exploited for lateral propagation purposes. Web servers are also commonly compromised to act as "watering holes", where an attacker will use their access to serve exploit code to visitors of corporate websites. Examination of log data relating to web servers can identify abnormal requests that might indicate attacker reconnaissance or exploitation activities, as well as providing insight into outbound requests that may show exfiltration activity from servers being used as data staging host.

Database Server

SQL injection (SQLi) attacks are a common method for exploiting weaknesses in web application software to obtain information from, or control over, a database server. Depending on the objectives of the attacker, SQLi attacks can be used to display the contents of protected database tables, add new user accounts to enable direct remote access, provide the ability to upload and execute attack tools and can even provide full remote desktop access to the server.

Database transaction logs can be used to determine what commands were sent to the database and whether or not they were successful. For attacks where the primary objective is to obtain the information stored within the databases themselves, log data can provide insight into what access the attacker has obtained and the impact of the activity.

If an intruder attempts to upload additional tools to the server via SQLi, database logs can potentially be used in conjunction with web server logs to extract and analyse the malicious payloads even if the original files have been deleted from the host.

Network

Authentication Servers

Authentication servers are used to authorise clients to join networks and access resources held on them. Although the implementations and the protocols used for authentication vary depending on the network environment, they typically log each user authentication attempt along with a set of metadata about the connection. Log data from authentication servers can be used during incident response to determine how an intruder is connecting to the network and to track their movements if they are propagating between hosts.

Control of an authentication server may represent a very attractive target to an intruder as they can potentially leverage this to have unrestricted access to a number of resources on the network. In addition to the logging of authentication attempts to the server, it is important to ensure that host logging is activated to detect direct compromise of the authentication server.

Firewalls

Firewalls can provide a rich source of data for network traffic types that contravene company policies based upon the rule sets they are configured with. Most devices can also track the state of traffic flows passing through them in order to perform content inspection of the sessions to detect, alert on and block malicious activity. The value of firewall log data will vary depending on where it is deployed within a network and the type of attack activity undertaken; firewalls on the border of a network may be able to detect incoming attacks and data exfiltration events, but will have no visibility of lateral movement between hosts.

Routers and Switches

Routers may be configured to permit or deny particular network traffic types and to log information about traffic that falls outside of these policies. In addition to this, many device models will log information regarding their system state and attempts to authenticate to their administrative interfaces. Some routers and switches have the ability to generate additional logging on network protocols that may be in use on the network. This information may be used in an intrusion detection capacity or as part of a proactive approach to improving network security by detecting implementation issues.

Hosts

Program Execution Auditing

Most operating systems provide some administrator-level logging of program execution, such as Audit Process Tracking within the Microsoft Windows environment. When enabled, this data can allow analysts to track when applications start and terminate as the host is running, potentially allowing them to identify the execution of malware

Process Execution Logging can allow analysts to track when applications start and terminate, potentially allowing them to identify the execution of malware

Other logging options can provide opportunities to detect malware persistence mechanisms as they can track services being started and scheduled jobs being registered. These log sources are best used in conjunction with an understanding of the typical configuration of hosts on the network so that anomalous behaviour can be identified.

File Access Auditing

The creation of log records relating to user and application access of sensitive files can provide an indication of potentially what data may have been lost in a network intrusion event. Within the Microsoft Windows environment, this is achieved through Windows Object Access Auditing, which logs accesses to files in administrator-defined folders and network shares.

In addition to file access logging within the operating system itself, Document Management Systems may also provide log data to track which documents are accessed by which users.

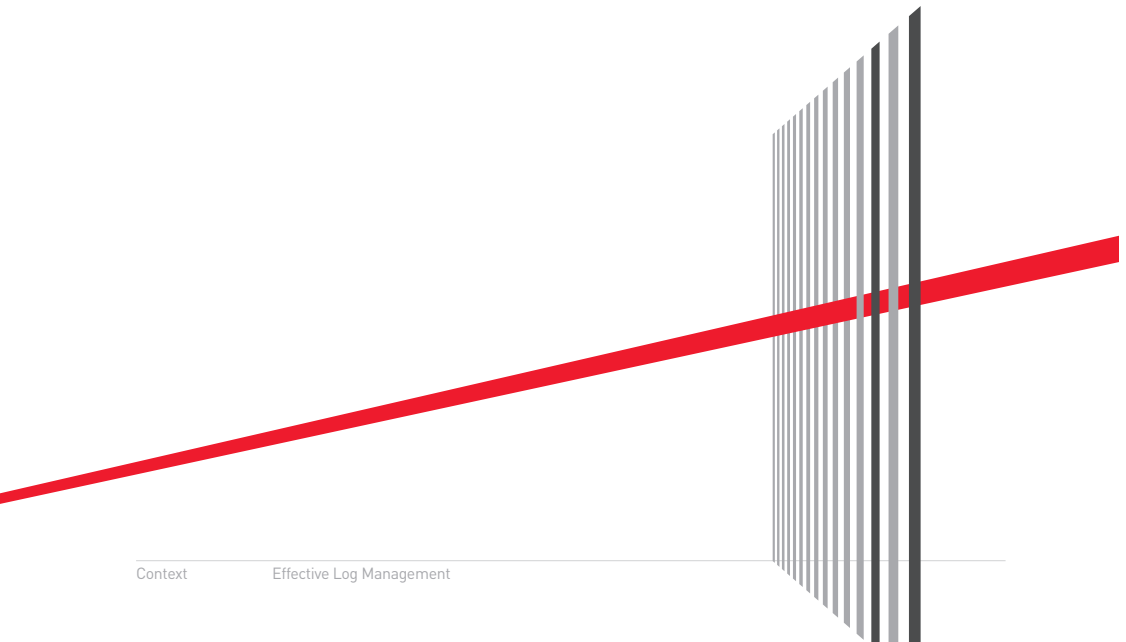
Security Software

Network Intrusion Detection Systems

Network Intrusion Detection Systems (NIDS) can record detailed information on suspicious network traffic flows based upon signature detection routines. Logging rates for NIDS will vary depending on the type and number of signatures deployed on the system and therefore the effectiveness of detection capability can be reviewed over time by examining log records to determine if the alerts are relevant and actionable.

Host Security Software

Security products such as anti-virus software are often deployed to provide a local level of protection to the host. Log records generated by this software can be useful for identifying intruder tools, although their value will strongly depend on the quality of signature sets and heuristic rules available to the application and whether the malware has been specifically designed to evade detection. Even if an attacker's malware is not detected by anti-virus applications, additional tools such as password dumpers may be identified through heuristic detection routines. In addition to the analysis of anti-virus detection logs, log files of the software itself can be useful as attackers may attempt to disable scanning routines, leaving evidence as log events



Remote Access

Virtual Private Networks

Virtual Private Networks (VPNs) provide a mechanism to remotely connect devices to a network and use associated resources as if they were connected to the corporate environment. While VPN technology is advantageous for connecting disparate computing resources and enabling remote working, the compromise of VPN access credentials can represent a significant vulnerability for a network as a remote attacker can interact with any host on the network that the user account has access to.

VPN logs can provide a good data set for behavioural analysis purposes and can be used to identify sessions that are markedly different from typical user interactions with the network.

Remote Desktop Services

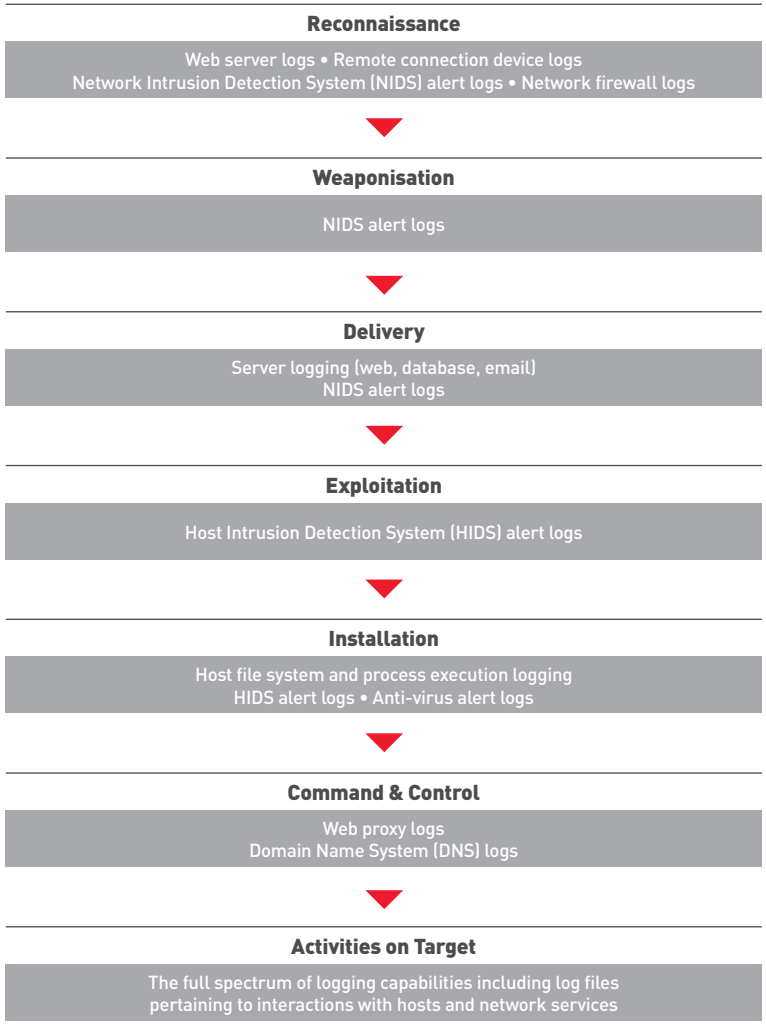
Remote desktop services allow users to connect remotely to machines within a network, either by establishing a new session on the host or taking control of an existing one, depending on the software used. While the remote host is not logically connected to the network as is the case with VPN connections, remote desktop services still provide many of the capabilities available to a local user such as access to network shares and applications installed on the machine. They also allow data to be sent between the remote host and the client through interactions between clipboards (copy and paste) and, depending on the software and configuration, "drag and drop" file transfer.

Remote desktop connections are often favoured by intruders as a mechanism for controlling a compromised host after valid user credentials have been obtained by malware or key loggers. Once connected via a remote desktop connection, attackers have the opportunity to use the graphical user interface of the host to install new tools, disable host-based security software and perform further reconnaissance of the network.

Analysis of log records pertaining to remote desktop connections can indicate which hosts within a network are controlled by an attacker based off known indicators of compromise, or can be used in a proactive manner to detect abnormal activity such as access to hosts outside of business hours.

The Kill Chain

The Kill Chain is a conceptual model used to define the stages through which an adversary has to progress to achieve an objective. When considering how log sources should be deployed and analysed, the stages of the Intrusion Kill Chain (Reference: http://papers.rohanamin.com/?p=15) can be a useful model for determining how a layered approach to log data generation can maximise the opportunities to detect an intruder through proactive analysis of a range of data sources.



Storage and Retention

Once the log source identification and selection process has been completed, the appropriate method for storing this information must be determined. There are three main log storage strategies that can be employed: local storage on each device, centralised storage on a log server or a combination of the two depending on the required detail and accessibility of log records.

Local logging distributes storage requirements across the enterprise as each device is responsible for supplying its own data store and therefore no separate database provision is required. Centralised logging can take the form of a networked file store that various agents can write their logs to, or a purpose-built solution in the form of a Security Information and Event Management (SIEM) tool.

Although some form of centralised logging is generally recommended as it ensures that larger amounts of log data can be pre-collected and correlated to enable real-time analysis for both intrusion detection and response purposes, it comes with a greater up-front provisioning requirement for primary storage and log backups. A combination of the two approaches may be considered as a method for overcoming many of these disadvantages at the expense of increasing overall administration overhead. In this case, the most important log sources and data fields are automatically forwarded to a centralised location for ingestion and real-time analysis, while more in-depth data is stored for shorter periods on individual devices and is interrogated on-demand when an incident is detected.

Suggested Log Record Retention Periods

Log Type	Retention Time	Notes
Proxy, web, authentication, remote access, firewall	2 years+	Longer retention periods are required for these high value data sources that can be used to identify the impact of new and existing persistent threats over typical attack timescales.
Email, IDS, anti-virus, database, network infrastructure	6 months – 1 year	While valuable, these data sources are often coupled with active alerting facilities that can notify security teams of incidents as they are detected.
Host process execution and file access	4 weeks – 6 months	Depending on usage, host logging can generate large numbers of events that are mainly used for verification after an intrusion has been discovered.

Security

However your organisation chooses to store log files, it is important to ensure that they are suitably protected from manipulation or destruction by a malicious party intent on covering the tracks of their intrusion. While early systems for logging events often did not consider intentional or inadvertent modification of log records as a primary consideration in their design, additional security measures are now appropriate as log stores present an attractive target for attackers.

Although the security capabilities vary between log management solutions, the following general recommendations can be made:

- **Ensure the integrity of the processes generating log records**
  Unauthorised users should not be permitted to modify or interrupt the processes that are used to create log files on devices. For example, users should not be able to turn off anti-virus software logging.
- **Limit access to local log files**
  If log data is stored on hosts, user accounts should not be given access to log files unless it is to append to existing records. If possible, no read, rename or delete privileges should be granted on log files.
- **Implement security on centralised log files**
  Centralised log stores can represent an attractive target for attackers wishing to modify log data across a range of devices. For log data that is sent to a centralised location, security measures such as digital signing or encryption can be explored to ensure the integrity of this data, although for large volumes this may have unacceptable efficiency or administrative overheads.
- **Secure log data transmission**
  If possible, any data transmitted between log sources and a centralised log store should be secured to prevent unauthorised modifications. While some devices and log agents support encrypted communication channels natively, other log sources that use plaintext protocols can be protected through the use of additional layers of encryption such as Internet Protocol Security (IPsec) tunnels. This may also be impractical for many organisations but is worth considering when high-value log sources have to transmit their data to centralised locations.

Utilising Log Data

Responding to Incidents

The role of log analysis in incident response involves the retrospective examination of log data following a security incident. Post-event filtering and analysis of these data sets using indicators of compromise can help develop a picture of an intruder's activity. This understanding can then be used to remediate the attack by identifying and cleaning the affected systems, determining and correcting the intrusion vector used and ensuring that no backdoors have been installed to maintain access to the network. As part of this, log analysis can be used to gain an understanding of the impact of the attack, including clarifying the origin and quantity of any data taken from the network.

The log analysis process during incident response will vary with the nature and extent of the attack. Often, a company will become aware of an incident through the detection of malicious activity by security software, a notification from an external organisation (private sector or government department) with wider knowledge of an attack campaign, or through proactive anomaly detection conducted by internal security teams.

In some situations, the scope and complexity of a security incident may warrant more in-depth analysis of log data that is beyond the resources of in-house network security teams. If this occurs, the expertise of an external incident response company may be required to assist in the handling of an intrusion event, during which time it is advantageous to have readily available and well organised log stores so that data can be quickly passed to incident handlers to enable rapid analysis and response.

Proactive Analysis for Intrusion Detection

Although log management is most commonly associated with incident response processes where log files are queried retrospectively based upon one or more indicators of compromise, log data can be used as part of an ongoing process of review and analysis wherever possible to identify attacks as they occur.

The most effective way to begin this process is to establish a baseline understanding of the typical activities on a network and to dedicate regular analyst resources to reviewing log data to identify anomalous events. The process to determine which events are anomalous will vary greatly between organisations and must be defined on a case-by-case basis by internal security teams or in consultation with external incident response providers.

Once a solid understanding of this baseline has been obtained, additional log sources can be introduced to provide a richer understanding of the state of the network. In addition, developments in analyst tradecraft such as heuristic identification of anomalous events (abnormal traffic volumes, remote connections from unusual locations or at unusual times, unexpected communication between hosts) will enhance an organisation's ability to proactively identify attacks against the network and limit any damage caused.

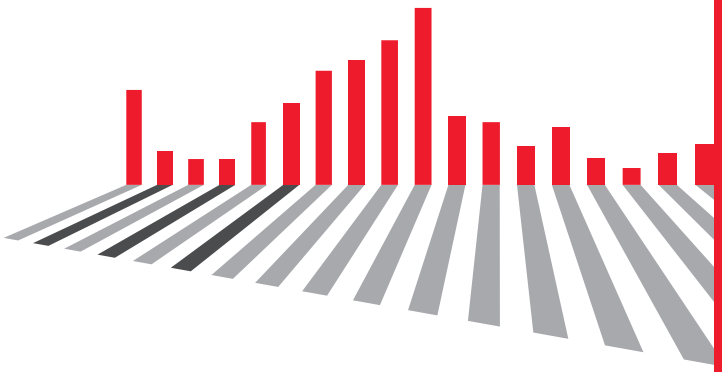
Utilising Log Data

Security Information and Event Management software

Security Information and Event Management (SIEM) software offers the capability to aggregate, store and display log data in a way that can be used to enable near real-time analysis of the security of a network from the perspective of each log source. If resourced and used appropriately, SIEM can help solve some of the most complex problems facing an organisation looking to improve the security of their network.

SIEM deployments generally consist of one or more database and log analysis servers and can acquire data either through agent-based or agentless mechanisms. While available products differ in the overall capabilities they offer, common functionality includes:

- **Log record aggregation**
  SIEM software provides a centralised location to aggregate logs from multiple data sources.
- **Data retention**
  Centralised storage also enables long-term retention of log data beyond limits that may be in place on each device.
- **Event correlation**
  SIEM software often includes the capability to integrate multiple log data sources and correlate events between them.
- **Visualisation dashboards**
  Many SIEM tools offer a range of customisable visualisation dashboards that can provide a high-level overview of the condition of the network and can highlight abnormalities in the data.



Case Study

In order to illustrate how an effective log management strategy can contribute to a successful incident response investigation, we present a case study where comprehensive logging enabled a detailed picture of the incident and a high level of post-investigation assurance that the incident was contained and remediated.

Company X received a victim notification issued through the Cyber Incident Response scheme informing them that a legitimate third-party website had been compromised and was serving malware (via a "watering hole" attack) from a suspected state-sponsored actor. There were indications that hosts on the network had been compromised after visiting the site.

Company X had a dedicated incident response and investigative capability and as such was able to stand up an internal team of significant size to conduct the investigation, with external experts providing oversight and direction.

Investigation via log analysis revealed that the attacker had not actively developed their foothold any further

During the investigation analysis of significant volumes (approximately nine billion records) of log data took place, primarily consisting of web proxy logs and host data logs. While the watering hole attack was found to have been successful and some malware was found to be active on the network, further investigation via log analysis revealed the attacker had not actively developed this foothold any further and there was no evidence of attacker activity elsewhere on the network. If the deployed malware had been activated by the attacker, the log store would have proved even more valuable as commands sent to the particular malware used in this attack could be decoded from data stored within proxy logs, allowing investigators to understand the attacker's activities more thoroughly.

As a result of the initial notification, the organisation was able to rapidly identify seven compromised hosts. Forensic analysis of the machines and reverse engineering of the identified malware gave a thorough understanding of the initial infection, provided further development of indicators such as IP addresses and traffic signatures, but not whether it had spread beyond those seven machines.