

Figure 2

Date	Time	Log_Name	Src_IP	Src_Port	Tgt_IP	Tgt_Port	Device_Type
21-Nov-05	12:10:29	Accept	192.168.65.65	1355	10.10.10.10	80	Check Point
21-Nov-05	12:10:29	List 102 permitted TCP	192.168.65.65	1355	10.10.10.10	80	Cisco Router
21-Nov-05	12:10:29	WEB-IIS ISAPI printer access	192.168.65.65	1355	10.10.10.10	80	Snort

mation for evidence admissibility.

To be creditable, electronic information has to be authentic, complete and trustworthy enough that it can be valued to a degree where it could influence the outcome of a legal proceeding. In short, there may be electronic information that is 100 percent admissible, but if the credibility is low, its integrity can be attacked and it will likely be excluded or at a minimum, its influence will be greatly diminished.

Normalization of Log Data

Because there are so many disparate forms of log data such as proprietary and legacy formats, XML, syslog, SNMP, ODBC and binary, the solutions used for collecting and analyzing logs will often have the capability to normalize the various formats into a common schema. This makes more advanced analysis – such as correlation, anomaly detection and pattern discovery – much more efficient and effective. In fact, NIST 800-92 states that, “To facilitate analysis of logs, organizations often need to implement automated methods of converting logs in different formats to a single standard format.”

As part of the normalization process, the logs need to be parsed. For mission-critical assets, 100 percent data capture

should be maintained. Depending on organizational policies, less critical assets may not require such holistic collection – thus reducing storage, processing and analysis resources. To get a better idea of normalization, consider the following figures.

Figure 1 (opposite page) illustrates multiple logs in their native format. These logs represent a remote printer buffer overflow exploit that connects to IIS servers over port 80 while crossing through a router and firewall.

Figure 2 (above) illustrates a subset of fields from logs in Figure 1 after normalization. Note that 100 percent of the data is retained and common fields between the disparate products and vendors become clear.

In addition to this normalization process, information can be reduced, filtered and aggregated to help manage high volumes of logs by minimizing duplications and/or unnecessary data. NIST supports this model by stating that, “Because of the volume of logs, it might be appropriate in some cases to reduce the logs by filtering out log entries that do not need to be archived.”

For assets that are not mission-critical, the flexibility of reduced field capture,

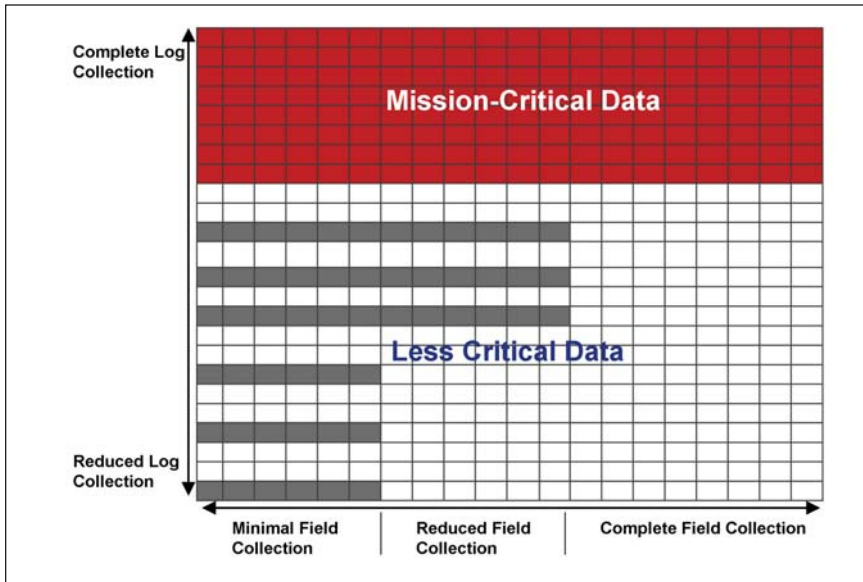
filtering and aggregation can save on bandwidth and storage requirements, as well as event processing and event analysis resources. Reduced field capture means that for less critical assets, every field doesn't have to be collected within a log – just the ones deemed most essential.

For mission-critical systems, log collection should be conducted on 100 percent of the fields within each log. Filtering will remove entire logs, not just fields within logs that are determined to be less essential. Finally, aggregation enables duplicate logs to be rendered as a single log.

Take for example a PING Flood. If a device receives 10,000 ICMP Echo Requests over just a few seconds from the same source, and the packets are identical excluding the time stamp, it would be advantageous to process, analyze and store a single event with a start and end time representing the length of the incident as well as a base count to represent how many packets were aggregated. Reduced field capture, filtering and aggregation should always be aligned with organizational policies.

Figure 3 (page 28) illustrates these concepts. Each small box represents fields

Figure 3



with an event log. At the top in red are mission-critical logs. Every field is captured and there is no log aggregation or filtering.

In the middle, some fields have been reduced within each log file as represented by sections of columns not being grayed out. Additionally, there has been some filtering and aggregation as represented by some rows not being grayed out. This depicts a situation where certain non-essential fields are removed and certain non-essential logs are being filtered and aggregated.

In the bottom section, field and log reduction follows to a policy that collects only the most critical fields, thus optimizing resource utilization. The important concept is flexibility. These capabilities enable an organization to have a much more customized approach to log collection and analysis – instead of a wholesale, boilerplate model where everything from critical business systems to printers, for example, are treated the same.

To put this into context, consider compliance regulations. Sarbanes-Oxley makes it clear that all financial transac-

tions need to be logged. However, if an organization is only interested in addressing this regulation, and only 10 percent of the logs on a database are related to financial transactions, why would they want to capture the other 90 percent if the logs don't make them more compliant, don't further mitigate risk, and simply don't provide them additional value – but instead will incur additional costs?

Laws and regulations are clear that authentic electronic information can be admitted as long as the information's integrity and accuracy meet given standards. Thus data that has gone through normalization, aggregation and so forth can be used as digital evidence, even when considering the concept of "original" information.

Original Information

Original information is the equivalent of having a piece of paper with an individual's inked signature on it – not a photocopy. For electronic information, the term "original" is difficult or even impossible to define. While the Federal

Rules of Evidence show a clear preference for original documents, for electronic evidence where duplicates are less of an issue because of the capability of making infinite perfect copies, various laws and regulations have responded. For example, the Electronic Signatures in Global and National Commerce (ESIGN) Act states that the three points to satisfy legal requirements for electronic evidence also satisfy the requirements of an original. The electronic evidence must:

1. Accurately reflect the information set forth in the contract or other record
2. Remains accessible for the period the law requires
3. Can be accurately reproduced in the future

With this understanding of the original, it is clear that for security logs, the emphasis needs to be placed on the nature and extent of the alteration that occurs. If the normalization and similar techniques maintain the meaning of the original, not only can it be used as litigation-quality data, but also it actually increases the usability and effectiveness of the log data.

Raw Log Data

Even with what has been addressed so far, any log collection and analysis solution needs to be flexible enough to support various organizational risk postures. If an organization feels that it wants to retain raw log data, above and beyond any legal requirements, it should certainly be able to, and do so in addition to the normalized information. It is not unusual to find an organization that approaches IT governance in a manner that goes beyond legal and regulatory requirements. However, with raw log data, the integrity of the asset that gener-

continued on page 30

ated the log can still come into question as easily as if the logs were normalized. A log collection and analysis solution is only as good as the data that enters it. Some assets, like mission-critical servers, may be secured with controls that mitigate the risk of information being altered, removed or added. However, systems like laptops will likely not have such stringent controls. The accuracy and authenticity of any log can be called into question, but logs from assets that do not have robust security controls may be more likely to get thrown out.

Author Authenticity

I'm including just a short reference to author authenticity. I've included this primarily because it helps to illustrate how easily logs can be overshadowed by legal issues. The Federal Rules of Evidence list criteria for questioning

electronic evidence include: determining if evidence is altered, if evidence is reliable, and the authenticity of the author. The author is a particularly interesting area to question. For example, even if malicious activity is tracked back to a person's home PC where he lives alone and there is copious supporting log information regarding his activities, if he has an unsecured wireless connection, he could simply claim that any malicious acts could have been perpetrated by anybody in his neighborhood and thus create doubt. This is yet another area that may be beyond the scope of a log management program.

Summary

Logs, regardless of being normalized or raw, need to ultimately be credible and admissible to be of any use in a legal situation. Also, the E-SIGN Act's

interpretation of the concept original information requires the logs to accurately reflect contracts or other records, remain accessible pursuant to legal requirements, and allow for accurate reproduction. The Federal Rules of Evidence state, if an organization relies on any record such as a security log in the routine course of business, that the information will be generally admissible. However, there is nothing to stop an expert from challenging its integrity. In the end, even with concrete log management, there are many intangibles beyond a log collection and analysis solution that are nested within the law; even the best evidence can be thrown out if it can't withstand scrutiny. ■■



Brian T. Contos, CISSP, is the Chief Security Officer of ArcSight.



ArcSight, a leader in Enterprise Security Management (ESM), provides real-time threat management and compliance reporting yielding actionable insights into your security data. By comprehensively collecting, analyzing and managing security data, ArcSight ESM™ enables enterprises, government organizations and managed security service providers to centrally manage information risk more efficiently. ArcSight's customer base includes leading worldwide companies across many verticals—and more than 20 of the largest U.S. federal agencies. For more information, visit www.arcsight.com.