



Using Log Files in Digital Forensics and Compliance

How to ensure that event logs are legally admissible and usable for investigations and compliance

This page intentionally left blank

Executive Summary

In the past few years, state and federal legislators and regulatory bodies have implemented a substantial number of new regulations designed to force companies to higher levels of accountability and information security standards. Some regulations, such as Sarbanes-Oxley focus on corporate accountability, but have major ramifications on how computer audit logs are handled. Other, industry-specific guidelines focusing on financial institutions provide very specific tasks relating to the collection, retention and review of logs from systems, applications and network devices.

As the importance of these logs has increased dramatically over this period, so has the confusion surrounding the legal issues of this data. On one hand, many believe that computer logs must be preserved in a pristine, unalterable format in order to be considered legally valid, while others believe that practical considerations allow a somewhat more flexible standard. Likewise, some professionals may claim that sampling or filtering of log records is an acceptable approach, while other evidence suggests that filtering data presents major obstacles to admissibility or credibility as evidence.

This paper draws upon published opinions as well as comparisons to other forms of evidentiary standards to present the argument that

- "Complete, accurate and verifiable" is the criteria that computer logs are to be held
- Filtering or sampling of log data is an unacceptable violation of this standard
- The preservation of log information is critical, not the format or file organization containing that information

Finally, this paper presents an overview of various common approaches to log management, along with an examination of the potential legal and technical obstacles that can arise with their implementations.

SenSage is a comprehensive purpose-built log management solution which provides the scalability, flexibility and redundancy to meet the challenges of collection and retention of the massive amounts of audit and system logs generated by enterprise applications, systems and network devices.



Introduction

Logs generated by IT system and network traffic activities are no longer simply of interest to IT managers and CIOs. Increasingly, logs have legal ramifications, especially in light of legislation like the Sarbanes-Oxley and the Gramm-Leach-Bliley Acts, but also in connection with financial damages involving identity theft, information privacy, hacking, and insider abuse. Because of their evidentiary value, logs must be managed as a legal record; they *must* be complete, accurate, and verifiable. Effective log management—collection, retention, presentation—is an essential tool for legal and fiduciary risk management.

Case #1: In the much-publicized trial of Martha Stewart, computer logs provided evidence of phone message alteration and the restoration of its original content—enough to tip off investigators and convince a jury of Ms. Stewart's attempt to mislead them. She was found guilty of obstruction of justice and could go to jail.

The result is that computer-generated logs—once a source of data that only the most die-hard techie could embrace—have emerged as one of the key chess pieces in legal risk management involving corporate security and liability. As the digital eyewitnesses to transaction and business processes, the digital equivalents of footprints, paper trails, and first-hand knowledge, logs are increasingly recognized as a critical asset for the board of directors, general counsel, internal and external auditors, and, ultimately, all C-level executives: the CIO, CFO, and the CEO.

What kind of enterprise needs to manage logs with government and regulatory mandates in mind? Actually, every enterprise does. Indeed, the standards and risks typically apply to *any* company that has electronic assets, which means everything from intellectual property and business records to personal information and communications with partners and customers. From the most macroscopic perspective possible, those companies need to ensure they've built *information assurance controls*—which translates into



Computer-generated logs—once a source of data that only the most die-hard techie could embrace—have emerged as one of the key chess pieces in legal risk management

Case #2: When the Securities and Exchange Commission was investigating Bank of America Securities for alleged illegal trading activity, the brokerage failed to furnish the appropriate records promptly—much of which was log evidence—and was found to have “willfully violated” the Exchange Act. The result: a \$10 million civil penalty—not because the SEC had actually uncovered illegal trading activity, but simply because the firm could not provide the sought-for data.

Computer logs, in other words, are no longer just troubleshooting tools for techies. They have major legal consequences for any enterprise who uses them—which is to say, of course, any enterprise at all.

Information security professionals have long been concerned about CIA—*confidentiality, integrity, and availability*—with regards to the data contained in computer networks. But the stakes have been raised dramatically as the Internet has emerged as a mainstream, highly accessible information resource and a substantial conduit for transactions of all kinds. In particular, the concern over privacy—highlighted by incidents of mishandling of private information and identity theft—has become a major issue in the minds of the public along with lawmakers and regulatory agencies. At the same time, corporate malfeasance (notably, the headline-making scandals of 2002) has led to new levels of scrutiny of corporate activity and new demands for provable compliance.

process, system, application, and network transaction logs.

The legal complexities associated with corporate responsibility and compliance are daunting, in part because the application and enforcement of recent legislation like the Sarbanes-Oxley Act and the Gramm-Leach-Bliley Act (GLBA) are still evolving. One thing is clear to the prudent executive, however, even at this nascent stage: to protect a corporation's interests and serve its business needs, logs must be managed appropriately. Which translates to computer logs must be *complete, accurate, and verifiable*. If logs are to provide the benefits of a digital eyewitness, their collection and storage should conform to the standards of legal admissibility and business reliability. Furthermore, mechanisms must be in place to allow for efficient analysis by systems administrators, law enforcement investigators, and regulatory compliance officers.

And make no mistake: The requirements for electronic data retention and controls are real. Compliance is critical, and the consequences of failing to address the letter of the law can be severe and will be enforced. Financial institutions, for example, found in non-compliance of the Gramm-Leach-Bliley Act can be subject to civil penalties of \$100,000 *per violation*. Officers and directors of that institution can be held personally liable as well, with penalties of \$10,000 *per violation*. Moreover, in a world in which privacy is no longer an abstract issue, the impact of non-compliance on an institution's reputation can be devastating, as several cases have already demonstrated.

Complete, Accurate, & Verifiable

Logs are the by-products of system and network transaction activities, and as such, they represent a large portion of the teeming ocean of data that engulfs the online/digital world. As a result, log management can make serious demands on a company's technological infrastructure. A substantial enterprise can easily generate terabytes of data in logs alone, pressuring IT administrators, legal counsel, and risk management teams to decide what to keep and how to manage the data deluge. These decisions, often guided by pressures of IT budgets alone, should be informed by the legal ramifications at the onset. Like the old adage about the cost of prevention, the cost of ignoring the law proves time and time again to far outweigh the technology expense at the front end. The elements common to these legal ramifications are **completeness, accuracy, and verifiability**, and these should inform legal risk management decisions.

Completeness: Even though the legal landscape surrounding logs is evolving, jurisprudence is steeped in precedent that favors evidence which paints a complete picture.

inadmissible. And incomplete logs can allow a regulator to draw inferences of malfeasance and noncompliance with an investigation. Having complete logs, on the other hand, will rebut any contention that only selective, biased data was gathered or that exculpatory evidence was not recorded.

Accuracy: For the same kinds of reasons, accuracy is a prerequisite for the successful use of logs in legal actions or in the context of compliance audits. Corporate due diligence and regulations like Sarbanes-Oxley are meant to ensure the accuracy of financial statements and the underlying IT controls. Accuracy of log data means that the time, date, and content of that log are the same as when it was created. Electronic copies are considered to be "best evidence" only if they accurately reflect the original.

Verifiability: As the recent accounting scandals have shown, credibility and integrity checks no longer rest on someone's word. If logs are to earn the labels of "complete" and "accurate" they must be verified as such. Some techniques such as "hashing" provide a digital fingerprint of logs that allows verification that log evidence is authentic



Absent accurate logs, companies risk not being able to rely on copies for proof purposes because legal standards require evidence to be "original."

Rationale #1: More often than not, the truth-value of evidence is only apparent when placed in context. To draw upon a familiar example: A video clip of a man being flogged by law enforcement might support a police brutality charge. But if that same tape, viewed in its entirety, showed that the man was holding a gun, the police response might well be justified.

Completeness in the context of logs means two things: individually, that activity is captured without gaps in time, and, collectively, that logs throughout an organization are maintained in the aggregate. With complete logs one can reconstruct the: who, what, when, where, why, and how of an activity involving a computer. Complete logs enable a "digital chain-of-custody," which mimics the court-tested method of proving that evidence is original and authentic. For example, maintenance of complete logs will ensure that events which look innocuous when viewed individually actually form a sequence of events that can prove an insider theft, an outside hacking attempt, or that a business transaction occurred.

Rationale #2: Completeness goes hand-in-hand with objectivity and reliability. A full and complete set of log data provides a truly objective picture of the digital landscape. This makes it possible for investigators, fact-finders, and even legal opponents to look at the data and reach the same conclusion, thus shifting proof issues elsewhere. Clearly, incomplete logs open the door to legal and regulatory challenge, allowing a counterparty to successfully argue that the logs are so unreliable as to be prejudiced and

days, months, or years later. But in addition, verification can occur by collecting logs from multiple sources and corroborating the content. Specifically, matching data in two separate logs from different devices verifies the accuracy of each independently. Other techniques used to enhance verifiability include the process of documentation each step of the log management process, creating a repeatable "digital chain of custody" and storing the data in multiple separate locations.

Access Turns Data into Information

If "step one" is to ensure that logs are complete, accurate, and verifiable, then "step two" is to figure out how to turn all that data into an information resource. It is essential to be able to extract information from the terabytes of log data an enterprise generates—quickly and in compliance with legal standards.

Presentation and analysis of logs is critical if they are to be "human readable" and useful in legal actions. Indeed, one of the key goals for today's enterprise is to manage legal risk and avoid legal costs, so self-policing and cooperation with enforcement officers and investigators is important. To do that effectively, logs must be reviewable and understandable. Access to compliance data and significant events, as well as disclosure of same, requires effective log analytics.

Sarbanes-Oxley, for instance, requires companies to disclose timely information to the public regarding material changes to the financial condition or operations of the company. Moreover, the Federal Trade Commission (FTC) maintains in its Safeguards Rules

(designed to help implement GLBA requirements) that it is critical to monitor, use, and review access records and logs.

Logs as Evidence

The corporate responsibility and privacy issues that are being addressed by legislation like Sarbanes-Oxley and GLBA are just two examples of how complete, reliable logs can be critical to a company. But because of their ability to function as evidence—as a sword or a shield—logs are often reviewed in connection with the discovery process of any kind of legal investigation. Just as a prosecuting attorney will scrutinize all aspects of an alleged crime, a forensic investigator will comb through logs during the process of discovery, looking for the digital equivalent of the smoking gun.

So for CIOs and general counsels, logs become part of the solution for managing legal risks connected to the control of information. Logs are increasingly used as “audit logs,” which are the primary evidence to demonstrate the reliability of

Logs and the “Internal” Hacker

One of the biggest information security problems doesn't come from hackers outside the enterprise but from “trusted” insiders within its virtual walls. Indeed, the FBI/CSI Computer Crime Study reported that more than half of those responding to a survey on the subject had experienced insider abuse of information resources. Obviously, logs tracing network traffic can provide evidence that an employee misused a corporate resource by, for example, sending confidential information to a competitor. That kind of activity can seriously damage a company's competitive position and lead to shareholder suits for negligence. Logs provide powerful, direct evidence that can prevent or mitigate the potential damage from this stealthy threat.

When someone is robbed, the victim is out of pocket for the stolen items. But when a public company is “robbed” of some of its information assets—even if the loss results in no direct impact on its financial health—additional costs in terms of liability must be considered. The victim-corporation can simultaneously be the perpetrator. In other words, a corporation can be sued for negligence if it cannot



Enterprises today should think of event logs as an asset. And like any asset, they should be managed accordingly: safeguarded against threats and collected and stored in a manner that adds value to a company's business by reducing legal risk.

electronic data and the processes used to create, manage, store, and protect digital information—including higher level financial data. Audit logs that are forensically sound can and should record *all* attempts to access or use corporate or customer data—and be easily accessible in the event of a legal action or inquiry. With “legally engineered” logs, companies can reduce the potential of losing a lawsuit, diminish the costs associated with discovery and defense, and increase the likelihood of forcing an opponent into settlement.

Logs can also be a resource to defend against actions related to corporate governance. In today's world, “due diligence” means that corporate managers have a duty to use what the law calls “reasonable care” to protect and preserve electronic assets—i.e., customer and transaction data—and have a well-thought-out plan to address threats to those assets. Logs can function as evidence that a company is effectively “self-reporting” on issues of security and privacy, that it is actively engaged in addressing reasonably foreseeable problems, and that it is able to cooperate with legal and regulatory bodies. This is only possible if logs are collected and stored appropriately so that they can be legally admissible evidence and provide the digital clues to answer critical questions concerning the reconstruction of any event that gives rise to a dispute. Complete and reliable logging technology ensures that companies can comply with any investigation.

demonstrate that it acted reasonably in protecting its information systems and data.

Consider a case of information theft that made headlines in March 2003, when hackers stole more than eight million VISA and MasterCard credit card numbers. The company whose job it was to securely protect those numbers could be found liable by customers or card issuers—if logs indicate that the data was easily accessible or that the company was on notice of previous real or attempted unauthorized accesses and failed to reasonably protect the target data. On the other hand, an accurate and complete record of logs that demonstrated multiple levels of protection and reasonable protection would help defend the company against a negligence action.

In addition to serving as buttress for the defense or for proving compliance, logs are an important weapon for companies bringing legal action against a hacker or information thief—because logs provide the digital evidence that can prove the legal concept of “malicious intent or knowledge.” Computer intruders often rely on the defense that their attempts to gain entry into a system are accidental—that they “didn't know they were doing something wrong.” But when logs show that an intruder made repeated attempts to access a server or data at, say, 1 a.m., the ignorance defense is exposed as a sham. Logs help establish accountability in a digital world which greatly facilitates responsibility-shifting and blame diversion.

Logs as an Asset

The protection of privacy and prevention of theft/misuse of personal information has become more than just a good idea—it has become the law of the land. And a robust, secure, and complete capability to gather and analyze logs is a vital prerequisite for complying with laws and policies and for protecting a company from civil lawsuits, regulatory censure or fines, and even criminal prosecution.

Enterprises today should think of event logs as an asset. And like any asset, they should be managed accordingly: safeguarded against threats *and* collected and stored in a manner that adds value to a company's business by reducing legal risk. The current business environment, characterized by corporate scandals and a rising tide of digital information and transactions, has made electronic audit information extremely important. Logs are controls inherent to computer systems, regardless of the size or scope of an enterprise. Because of their ubiquity, logs offer a valuable tool to demonstrate the quality and integrity of information and processes to those inside

While Sarbanes-Oxley applies to all public companies, the Gramm-Leach-Bliley Act (GLBA), made effective in 2002, is more narrowly focused on financial services firms. Its mandate is clear: It requires the secure storage, transmission, and disposal of covered electronic records to ensure data integrity, protect consumers' privacy, and prevent identity theft. Banks, brokerages, and other financial institutions must use appropriate oversight and audit procedures to detect access by a hacker or other intruder or the disclosure or theft of customer information.

How do computer-generated logs relate to human-generated legislation like GLBA? Logs provide a bridge – they are the technical mechanisms that enable compliance with the requirements defined by law. The Federal Trade Commission issued a set of Safeguards Rules to help institutions implement GLBA requirements. The Rules' underlying premise is that the ability to monitor, use, and review access records and logs is critical for compliance. In addition, the Federal Financial Institutions Examination Council (FFIEC) has weighed in on the privacy/intrusion issue with its "Interagency Guidance on Response Programs for Unauthorized Access to Customer Information and Customer Notice," which indicates that logs are a primary component of



The protection of privacy and prevention of theft/misuse of personal information has become more than just a good idea—it has become the law of the land

the enterprise as well as the legal and regulatory bodies that oversee the enterprise.

Logs and recent legislation: Sarbanes-Oxley and Gramm-Leach-Bliley

The Enron scandal became the poster-child for poor corporate governance and one of the driving forces behind the passage of the Sarbanes-Oxley Act of 2002, which requires that public companies must maintain, review, and certify internal controls regarding financial information. While this is commonly thought to apply primarily to **audited** financial statements and accounting standards, Sarbanes-Oxley also applies to the technology controls involved in the prevention and detection of fraud.

Logs have a vital role to play with respect to Sarbanes-Oxley. Beside commanding certification of internal controls, the Act requires companies to inform the public in a timely fashion about *any* material changes to their financial condition or operations. A computer or data security issue can easily rise to the level of "material"—just ask the victims of malicious worms, as attested by their million dollar assessments of resulting damages. By providing the evidence of hacking attempt or data theft, logs can provide a signal to the general counsel and CFO, possibly triggering a press release to disclose the issue.

fulfilling security and notification provisions.

Log Management: Options and Issues

Commercial software vendors and in-house developers have both attempted to meet the challenges of log management by implementing common approaches.

Option 1: Manual Collection and Review

Even now, a surprising number of enterprises continue to perform log review of mission-critical applications and systems in a decentralized, ad hoc manner. In some such situations, organizations tend to lack a central policy or strategy for regular review of audit trails and other system logs. In other situations, for example large government organizations, the availability of personnel make scaling through manpower a viable alternative to scaling through software.

The potential problems with this approach are numerous: it is error-prone, manpower intensive, and provides little or no ability to identify incidents or trends by corroborating records from disparate files. Perhaps the most significant weakness of this approach is that any ad hoc or informal approach to log review will be subject to rigorous scrutiny, and possibly ruled inadmissible, in a legal challenge.

For these and other reasons, most firms with more than a few systems and have any in-house programming talent have moved on to the second option.

Option 2: In-house Development and Log Consolidation

By far, the most common option is often built upon in-house developed utilities created for system management and troubleshooting. Common approaches include creating central syslog servers, extensions to log rotate scripts, and command-line or web CGI utilities to perform queries against the data and generate reports.

Enterprises using this option quickly realize the limitations of working with raw logs. Increasingly voluminous and disparate files create management challenges: compressing files to save space leads to substantial decompression penalties, and correlating information from different log types requires complex parsing of records during queries. As a result, historical analysis and investigations become impractical if not impossible to perform in reasonable time periods. Moreover, this approach requires that in-house developers acquire the subject matter expertise to interpret the underlying log files in a meaningful way.

is properly tuned and indexed for the anticipated queries. The overhead associated with queries which cannot take advantage of an index (i.e. results in a table scan), however, can be prohibitively slow or exceed available system resources.

In fact, a number of performance issues arise when using relational databases in high-volume log management architecture; many of these issues are directly related to how the database is indexed:

- Insertion Performance: indexing impairs insertion rates, sometimes to below the rate at which they are generated
- Query Performance: queries not leveraging an index result in poor table scan performance
- Disk Usage: disk consumption is inflated by indexing; combined with temporary and rollback space allocation, required disk space may exceed 4 times that the actual data size
- Index Degradation: high insertion volumes, combined with frequent deletes, lead to degradation of the efficiency of the indices, which must be periodically rebuilt



Challenges with homegrown systems...compressing files to save space leads to substantial decompression penalties, and correlating information from different log types requires complex parsing of records during queries.

From a legal perspective, organizations adopting this option must also contend with the issue of demonstrating the authenticity of log file information. Data in file-based storage can easily be modified by malicious individuals with system administrator rights. To counteract such a scenario, file-based systems must incorporate a secondary, more secure, storage solution and/or mechanisms such as cryptographic hashing to ensure data integrity.

Finally, organizations implementing home-grown systems may be subject to more arduous proof of the reliability and accuracy of their systems, compared to organizations adopting commercial products whose reliability has been established.

Option 3: In-house or Commercial Products built on Legacy Systems

Another option adopted by both commercial and in-house developer centers on storing log data in a relational database. Each record is broken out into specific fields stored as columns. This approach often includes a data "normalization" process, i.e. storing disparate log types into a common schema.

This option provides flexibility in constructing queries for log data investigation and analysis – as long as the database

Organizations attempting enterprise log management with RDBMS-based products frequently discover that the solution fails to handle the volumes of logs generated by the enterprise. It also fails to retain the data over sufficient time periods. As a result, log files are filtered to include only what the vendor considers being "events of interest", and data must be purged after relatively short periods. Additionally, some vendors choose to not preserve all information from log records in order to save space and improve performance. This type of filtering risks omitting valuable information as well as violating the completeness criteria.

The legal effect of filtering and purging is twofold: the resulting data may be incomplete and cannot present an unbiased view of a sequence of events, a situation which can affect its evidentiary admissibility or weight. If the data is admitted, however, the credibility that a factfinder attaches to the data may be weakened without the corroborating data that was selectively filtered. Therefore, RDBMS log management systems can pose evidentiary challenges by impacting the admissibility criteria of completeness and accuracy.

Because of issues associated with the use of RDBMS systems for long term/high-volume log management, a number of vendors have begun to develop alternative or hybrid systems attempting to provide the flexibility of SQL with storage arrangements that are better suited to log data.

Option 3a: Raw Logs with Relational Indexing

In this option, log data is stored in flat text files (as in option #1), while a relational database is used to store indexing information and other meta-data about the files in order to facilitate locating specific records within the managed files.

The primary, if not sole, advantage of this approach is that it preserves the log files in their original format instead of a reorganized logical database format. In the past, there was ambiguity in interpreting what was an "original" for purposes of meeting the Best Evidence Rule for admissibility, with some legal authority opining that only raw logs would suffice. However, logs that are 'shown to reflect the data accurately' sufficiently meet this standard [Fed. R. Evid. 1001(3)]. For reasons of practicality, courts have recognized the importance of the informational substance in the logs rather than the format, and have focused on the completeness, accuracy, and verifiable criteria instead. As a result, the burden of preserving files in their raw, unadulterated format is not a prerequisite to admissibility.

Apart from the dubious advantage of file preservation, this option suffers from many of the issues associated with the first two options. Most of the problems with relational databases

stem from the volume of inserts and deletions, and their effect on the related indices; storing actual records outside the database will reduce the disk space consumed by the database, but will not solve index-related issues. Furthermore, queries against data that cannot leverage indexing data stored in the database result in the poor performance associated with flat file-based systems.

Option 3b: Two-tier Storage with On-Demand Loading

Another hybrid option implemented, for example, in at least two Security Information Management (SIM) products involves storing logs in a two-tiered system in which data in the top tier is accessible, whereas data in the second tier must be migrated into the top tier on demand when needed. In these implementations, the top tier is a relational database, whereas the second tier consists of compressed files.

This approach has the advantage of using the database for all queries, avoiding the need to search through flat files. It suffers, however, from the substantial overhead involved in identifying and loading the appropriate second-tier data when required, as well as purging unneeded data from the top tier in order to keep that database to a reasonable size.



The SenSage Advantage

SenSage's Enterprise Security Analytics overcomes event-data management obstacles and limitations of RDBMS-reliant log management systems. SenSage provides the most scalable means to centrally aggregate, efficiently analyze, dynamically monitor and cost-effectively store high-volumes of event log data while preserving chain-of-custody and streamlining forensic investigations.

Without reliance upon agents, SenSage collects data from a broad range of sources at a sustained rate of 87,500 events per second, and scans that data at 2,000,000 records per second (assuming a five-node cluster). This real-world scan rate is that of a complex sub-string URL search with full data extraction. This allows audit and IT staff to get faster results from a huge repository of multiple-source event data. When storing event data, SenSage can compress it to as little as 10 percent of the original, raw, event-log data; and 2.5 percent the size of that stored using an RDBMS. Since SenSage is based on purpose-built repository with self-optimized storage, it also avoids RDBMS tuning and archiving. Clearly, users can store all their event data without outrageous storage and administrative costs.

The compressed data is organized in a way that supports very fast queries, especially the iterative type, pattern-matching and sub-string searches so characteristic of investigations. As a result, digital investigators can efficiently analyze all the event data in the repository, as opposed to using data subsets or restoring archives. Furthermore, SQL queries can be executed against the repository, supporting the ad-hoc nature of investigative activity.

Data integrity is preserved based on the strict implementation of security controls throughout the collection and storage process. SenSage's log adapters and data mapping schema ensure that semantically complete log records are stored. Additionally, the SenSage Collector maintains a complete audit trail, from log file retrieval through loading. This audit trail is automatically loaded and stored in SenSage's Scalable Log Server (SLS). The Scalable Log Server is built on a secure configuration that protects against illicit manipulation of the data once loaded. The SLS uses encrypted transport and is highly redundant, powered by tamper-resistant database technology. Furthermore, due to the following characteristics, file-level tampering is virtually impossible:

- 1) The SLS stores copies of each log file generated on two separate servers, and
- 2) The SLS can be configured to store a signed hash of each log record in an additional column.

Because the SLS does not support SQL UPDATE or DELETE, it is resistant to SQL injection and other common SQL-based attacks. All report viewing is secured via SSL.

Summary

SenSage was architected to address the specific problems of audit data collection, retention and analysis. Designed to provide the high performance, scalability and compression required for large volume log management and compliance needs, SenSage provides a unique solution. The following table summarizes the varying benefits of different log management approaches, and highlights SenSage's value proposition for forensic investigations and regulatory compliance.

<i>Benefit</i>	<i>Option 1 Decentralized</i>	<i>Option 2 In-House Scripts</i>	<i>Option 3 Legacy Systems</i>	<i>SenSage</i>
Completeness	No – silos of information make completeness unverifiable	Yes	No – database performance issues force users to filter data	Yes Scalability allows management of all logs with long-term retention
Verifiability	No – no documented process	Maybe – lack of a commercial vendor puts burden of proof on the customer	Maybe – depends on specific vendor	Yes Digital chain of custody demonstrates repeatable process
Accessibility	No – data is spread across multiple locations making correlation impossible	Limited – complex queries against unstructured data can be complex to write and time consuming to run	Limited – structured data is queryable, but only efficient when queries correspond to the existing indices	Yes Data is structured, with exclusionary indexing and parallel processing providing performance
Scalability	Yes, but data is spread across a variety of systems and requires additional manpower to process	No – unprocessed data cannot be easily searched in large volumes	No – underlying database systems create bottleneck	Yes Cluster architecture provides parallel processing of loads and queries
High Availability	No	No, unless developed in-house	Maybe, depending on complexity of underlying RDBMS	Yes Cluster architecture provides redundancy and automatic failover
High Compression	No	No, unless developed in-house, with corresponding query performance degradation	No – index, rollback, and temporary space allocations cause data to grow, rather than compress	Yes Unique storage model provides highly efficient compression ratios and decompression performance
Tamper-resistant	No – files can be easily modified	No – files can be easily modified	No – modification through “UPDATE” or “DELETE” commands is trivial	Yes Data is distributed across the cluster with automatic secondary copies – “UPDATE” & “DELETE” are disabled

About the Author

Erin Kenneally, M.F.S., J.D.

Forensic Analyst and Attorney with the San Diego Supercomputer Center, University of California at San Diego



Erin Kenneally is a Forensic Analyst and Attorney with the University of California San Diego Supercomputer Center, where she researches and consults on criminal and civil cyberlaw issues related to computer forensics, information security, digital evidence and privacy technology. Kenneally is active in the High Technology Criminal Investigations Association, the American Bar Association's Science and Technology Information Security Committee, and the Global Privacy and Information Quality Working Group.

About SenSage, Inc.

SenSage, the leading provider of enterprise security analytics, offers unparalleled performance and a scalable means for organizations to centrally aggregate, efficiently analyze, dynamically monitor and cost-effectively store massive volumes of event log data. Our solutions empower companies to readily respond to business-critical threats, conduct thorough and precise investigations, and maintain compliant operations. Based in San Francisco, CA, SenSage currently protects Global 2000 customers in financial services, government, healthcare, manufacturing, and technology. The company markets its product directly and through partners including Cerner, EMC, Hewlett-Packard, and Lockheed Martin. For more information, visit www.sensage.com.

The information provided in this white paper is for guidance purposes only and is published as legal analysis, not legal advice. While every effort is made to provide quality legal information, there are no claims, promises or guarantees in respect of any specific legal problem. As legal information must be tailored to the specific circumstances of each case, and laws are constantly changing, we recommend you consult a lawyer if you want professional assurance that our information, and your interpretation of it, is appropriate to your particular situation.



SenSage, Inc. • 415.808.5900 • www.sensage.com

©Copyright 2005, SenSage, Inc. All rights reserved.

SenSage and Scalable Log Server are trademarks of SenSage, Inc. in the United States.