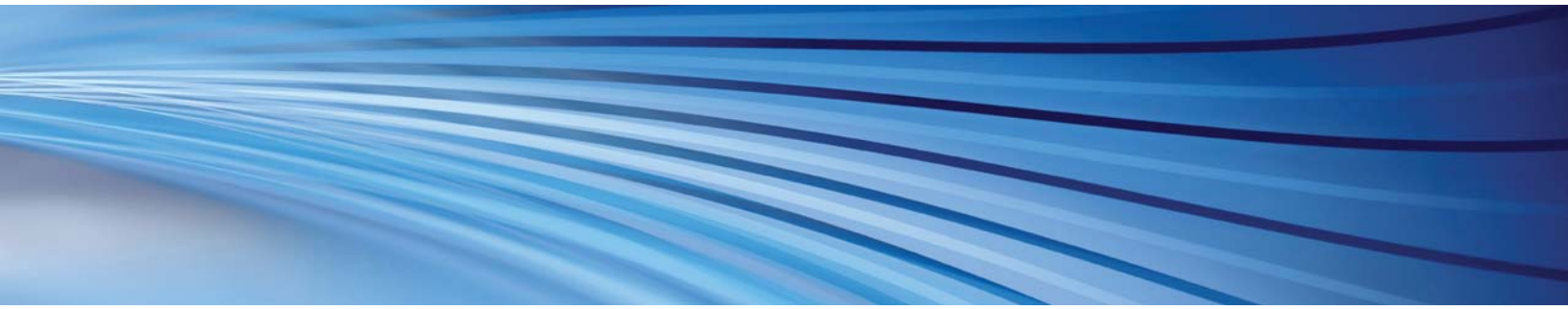
White Paper

Why Correlation is Critical for Best-Practice SIEM



Recently, one of our retail customers reported that they are running upwards of 200 million messages per day through the CorreLog Enterprise Server, just from the IBM z/OS mainframe. Collecting all of this data is certainly a necessity for compliance standards, forensic analysis and managing end-user performance and availability. But how can analysts possibly make sense of all the data filing through every minute. Collecting the data is only part of the equation and research suggests log message collecting is becoming more commonplace in today's IT shop. According to the latest SANS Institute Log Management Survey Report, 89 percent of respondents indicated they collect log data, a significant improvement over the 43 percent who responded to the first SANS survey just 7 years ago. However, organizations employing correlation for automated threat detection and the leveraging of log data for managing corporate IT security and compliance are patchy at best.

Today, many organizations are turning to correlation which helps bring meaning to the massive amount of data collected. These correlation engines are able to deal with massive log message volumes from many diverse sources. Those sources might include Syslog messages, SNMP traps, Windows events, security and application logs, firewalls, routers and network hardware servers and applications, and many others. If the software vendor can correlate this data in real time, it might mean the difference between staving off a threat or being victimized by it.

Additionally, IT departments are tasked with the need to capture, and store these messages for auditing and compliance. Once these messages are captured they must be reviewed for internal security threats as well as mandated regulatory compliance in the form of Sarbanes-Oxley, PCI DSS, HIPAA, FISMA and many others.

As "doing more with less" becomes the battle cry for IT, what organization has the time and resources to look at all the data for patterns and relationships that may signify security threats or performance issues? An ideal scenario would be a systematic approach to identifying relationships (correlating!) between log data that could indicate potential problems as they occur. With millions of messages collected daily in a single IT environment, the impossibility of a handful of IT admins pouring through all of the data to identify compliance issues and security threats is a dark reality.

So how do you leverage this mass of log data into decision support and actionable tasks to combat security threats and compliance breakdowns? This paper provides insight into CorreLog's approach to collecting, tracking, searching, and correlating log data for security information and event management (SIEM) automation.

First, we will review a time-tested approach for log file reception and aggregation. Next, we will study how CorreLog identifies patterns across seemingly unrelated log data that correlates to potential security threats. And finally, we will review a proactive approach to managing threats that are automatically generated from the CorreLog server.

Log Data Management

One possibility to deal with the chronic problems associated with enterprise-scale data aggregation is through a distributed management approach, where (in a large environment with millions of nodes) multiple copies of the server exist within an environment, executing as management agents. CorreLog Security Correlation Server manages this issue by collecting and analyzing data from a specific locale and then reporting the data (via a "ticket" system) to a higher-level data collector or

aggregator. In addition to reducing overall network traffic through a central aggregator, this distributed approach provides a more secure transaction by keeping the data close to the original source, and not transmitting the data any farther than it has to go.

With millions of messages – many of them meaningless – coming through the system, bandwidth integrity is obviously at risk. CorreLog addresses this problem by focusing on a two-tier architecture, where multiple copies of CorreLog Server gather information, save it, reduce it, and send only pertinent notifications to a higher level system. In this approach, CorreLog churns at an events-per-second rate of more than five million messages, with the capability to manage more than one million devices.

For compliance and auditing, CorreLog compresses and archives data, retaining it for a period of more than 10 years (5,000 days) and generates archival data such as MD5 checksums and Security Codes. CorreLog can collect in excess of 100 Gigabytes of data each day at a single site, and save this data online for up to 500 days (given enough storage).

To support this role as a data collector and real-time correlation agent, the CorreLog Server incorporates an SNMP agent (implementing secure SNMPv3) to allow remote management of the program, and implements a "CorreLog-MIB" that contains functions to manage message rates and initiate high-speed searches of data across the entire enterprise. In a complementary role, the CorreLog Server coexists with other processes on resident platforms, therefore making installation relatively non-intrusive. This translates to a simpler installation and ease of maintenance at multiple locations, local or remote.

These features provide a somewhat different approach to aggregating data from what is commonly found in other SIEM managers. CorreLog is ideally suited to work as a complementary component of a larger security management strategy, where the CorreLog Server operates as both "management agent" and "server" program. This approach enhances real-time security, reduces bandwidth, and provides a method of scalability to securely manage millions of events per second, and millions of devices across a large enterprise.

Correlating Log Data for Actionable Information

CorreLog uses a variety of exclusive correlation techniques that decode meaning from large numbers of received messages. The solution incorporates high-speed, index-driven search at its front end, and employs Artificial Intelligence technology at its back end, creating an advanced correlation engine with the ability to perform semantic analysis of messages in real-time. The system utilizes correlation threads, correlation counters, correlation alerts, and correlation triggers, which refine and reduce incoming messages into data that is easier to make sense of and more importantly, linkable to other messages that collectively could indicate a security breach. The search engine is interactive and permits fast searching of terabytes worth of data, while the correlation component reduces the enormous amount of data into brief and meaningful incident reports that are auto-generated.

At the heart of the CorreLog server is a unique correlation system that reduces a stream of incoming messages from various different device types into a series of actionable "tickets" that can perform automated actions and/or make suggestions to human operators on how to handle and mitigate threats.

CorreLog uses a unique type of correlation, referred to as "semantic correlation," which looks for meaning in messages or combinations of messages. Incoming messages are translated into information that is meaningful to organizational security and actionable to stakeholders tasked with threat management and compliance. Taxonomy and categorization of data is at the center of this system. The system automatically catalogs information by IP address, username, facility, and severity, as well as arbitrary keywords, regular expressions, logical operators, global variables, and macro definitions. This information can be qualified by the time of day and/or by preceding messages.

Any system that monitors organizational security and compliance should ultimately operate with continuous improvement as a standard practice. The CorreLog Server operates as a "recurrent" neural network, meaning that all of its output is transferred back into the input of the system, making it self-aware. Back-propagation and training is built into the system through an "auto-learn" function that automatically adjusts thresholds of alerts based upon message rates and their standard deviation values (see Figure 1). Specifically:

1. Input messages are compared to match patterns, and are threaded in conjunction with triggers, creating catalogs of messages. This cataloged view provides the user message visibility at any stage of correlation.
2. Each trigger and thread combination maintains a count of messages. The counter rates can be detected by the alerting component, which compares the counter rates (over an interval of time) to one or more thresholds.
3. When threshold counts are breached, the alert component generates more messages, and these messages can be further correlated with additional triggers and threads, which can generate additional threat detection.
4. Multiple stages can be created, where messages are "bussed" into all triggers and threads to create a network of correlation rules. The output of each stage is made available as a possible input to all other stages in the system correlating groups of messages to a higher level of threat detection.

The specific thresholds and connections between each stage of correlation define the types of patterns that are matched. At the final output stage (which can be the first stage, or a much later correlation phase) the highly reduced messages can trigger actions, or open action tickets.

One of the characteristics of this correlation arrangement is that it is well suited for filtering out false positives. This is often cited as one of the most unique and useful aspects of neural network architecture. Neural networks are highly adept at matching patterns in "noisy" environments, because each stage serves to reduce noise and eliminate extraneous false-positives. To increase this filtering action, the user can configure more stages to the neural network.

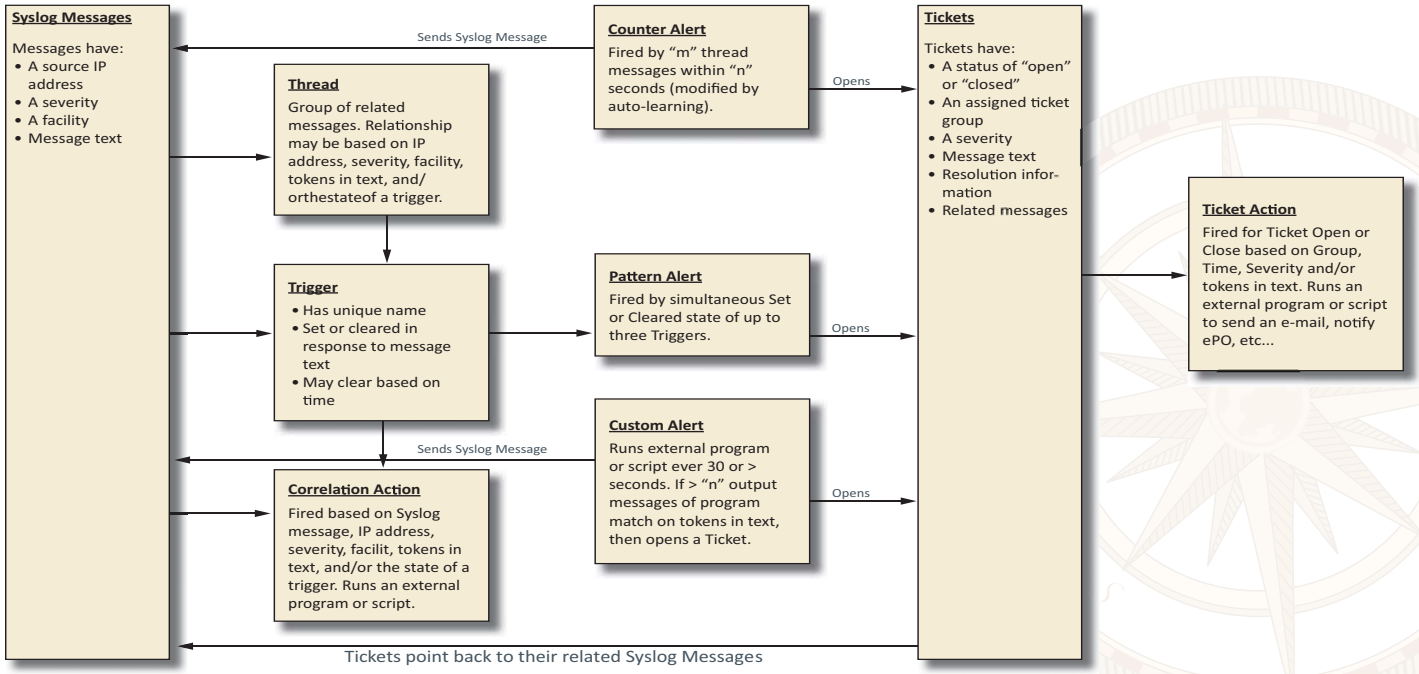


Figure 1

Turning Information into Action from the Correlation of Message Log Data

Collecting log data and presenting that data in a single, consolidated view is not revolutionary. However, the ability to take raw log data from disparate sources and apply logical correlation rules to that data is an approach unique to only a handful of SIEM solutions. CorreLog's proactive correlation technology uses threads to send alerts and can automatically open help-desk tickets with most help desk systems. The technology can also take action based on security or regulatory compliance parameters.

A "resolution ticket" attribute could contain the precise security event that has occurred, including the various messages that caused the ticket to be opened. Both a human operator and a software system can act upon the ticket (See diagram in Figure 1). Tickets can be sent to third-party incident management systems (to enforce a workflow) or can be relayed to another SIEM system, as is the case with CorreLog and McAfee ePO. CorreLog is a Certified McAfee partner.

The system also incorporates a simple and extensible "actions" capability, which permits the user to target specific messages based on device, keyword, facility, severity and/or time of day, and can then run programs on that data. It includes utility programs to update relational ODBC databases, relay Syslog messages, send SNMP traps, send e-mail, and perform other actions linked to the creation of a ticket. The facility is designed for easy extensibility by administrators and developers to extend correlation and ticketing services of the program to additional complementary system resources.

Another key capability for best-practice correlation of log message data is high-speed indexing. The backbone and core to making sense of the massive amount of log data is dependent on the ability to index hundreds of millions of events based on keyword searches. This is similar to an Internet search where the results come back instantly. There is no database required, nor detailed search parameters. This instantaneous search method employed by CorreLog allows for real-time correlation threads to execute rules on message data in real time, as it comes into the CorreLog server.

Conclusion

We've seen a rise in log data collection over the past 10 years. However, collecting the data is only a part of the equation. Our experience at CorreLog shows that the best success for customers is centered on their ability to derive meaning and subsequently automated actions from the millions of log messages collected daily. With a systematic and practical approach to correlating messages into real-time alerts, organizational security and compliance are enhanced, and with little added burden to existing IT human resource. In correlating seemingly unrelated log data into potential threats across a single desktop user interface, CorreLog leverages existing data through existing systems resulting in a best-practice approach to managing log data for SIEM.

About CorreLog, Inc.

CorreLog, Inc. delivers security information and event management (SIEM) combined with deep correlation functions. CorreLog's flagship product, the CorreLog Security Correlation Server, combines log management, Syslog, Syslog-NG, SNMP, auto-learning functions, neural network technology, proprietary semantic correlation techniques and highly interoperable ticketing and reporting functions into a unique security solution. CorreLog furnishes an essential viewpoint on the activity of users, devices, and applications to proactively meet regulatory requirements, and provide verifiable information security. CorreLog automatically identifies and responds to network attacks, suspicious behavior and policy violations by collecting, indexing and correlating user activity and event data to pinpoint security threats, allowing organizations to respond quickly to compliance violations, policy breaches, cyber attacks and insider threats. CorreLog provides auditing and forensic capabilities for organizations concerned with meeting SIEM requirements set forth by PCI/DSS, HIPAA, SOX, FISMA, GLBA, NCUA, and others. Maximize the efficiency of existing compliance tools through CorreLog's investigative prowess and detailed, automated compliance reporting. CorreLog markets its solutions directly and through partners.
Visit www.correlog.com for more information.

CorreLog, Inc. • 311 Conners Ave. • Naples, Florida 34108 • 1-877-CorreLog • 239-514-3331 • info@correlog.com