

Heuristically Advanced Warning Konsole, “HAWK,” consists of a multi-component solution for the handling and data mining of multiple third party alerting sources.

Abstract: *Faced with a growing backlog of event-log entries and false positives, businesses are turning to event-log correlation tools to streamline security analysis and performance monitoring.*

The Heuristically Advanced Warning Konsole, (HAWK), consists of a multi-component solution for the handling and data mining of multiple third party alerting sources. This functionally provides an analyst or administrator the ability to consolidate and correlate information into grouped sets, enabling the user to monitor for intrusion anomalies and/or otherwise critical information related to your network infrastructure.

HAWK uses data retained from external alert sources to logically create associations between similar data. The similarities are drawn by utilizing data-mining and analysis in order to correlate information for secure storage and future retrieval. HAWK logs the correlated data into a relational database available for advanced data management.

*By William M. Townsend
Managing Director
Tel. 512-228-2400
E-mail: Townsend@interminds.com*

Highlights

- Fully automated in-network technologies are critical to providing behavior changing service offerings.
- 63% of IT professionals and 55% of Global 2000 IT managers are not satisfied with their current log file analysis process.
- 89% of IT professionals would welcome an all-in-one analysis platform.

Publication Date: May 24, 2008

Introduction

The patent-pending Heuristically Advanced Warning Konsole, (HAWK), consists of a multi-component solution for the handling and data mining of multiple third party alerting sources. This functionally provides an analyst or administrator the ability to consolidate and correlate information into grouped sets, enabling the user to monitor for intrusion anomalies and/or otherwise critical information related to your network infrastructure.

HAWK uses data retained from external alert sources to logically create associations between similar data. The similarities are drawn by utilizing data-mining and analysis in order to correlate information for secure storage and future retrieval. HAWK logs the correlated data into a relational database advanced data management.

HAWK's unique differentiators include Naïve Bayesian histogram analysis, standard deviation analysis, K-Mean historical analysis, and statistical trend analysis.

HAWK provides advanced log correlation for many resources and appliances, including:

- Unix/Linux/Solaris Security Logs
- Microsoft Windows NT 4.0/2000/XP/2003 and Vista:
 - Client agent correlation of Active Directory, IIS, Exchange, Windows Security Events and more.
- Routers, Switches, Firewalls
 - Cisco Network Appliances, FWSM/PIX/ASA Firewalls
 - Checkpoint Firewalls
- IBM AS/400

The HAWK Correlation Solution consists of several separate components:

- HAWK Pulse (**HCPULSE**) - capable of loading dynamic plug-ins supporting a myriad of vendors in order to consistently “pulse” for unique data on remote vendor systems. Vendors include Cisco NIDS, SNMP (v1/v2/v3), Snort IDS, and more.
- HAWK Syslogd (**HCSLOGD**) - capable of correlating and filtering syslog messages from its myriad of vendor plug-ins with support spanning from Solaris/Linux servers, Novell, Cisco, Juniper, and more.
- The HAWK Information Event Konsole (**IEK**) acts as the management and data retrieval interface with the relational database. The IEK is the primary method of interacting with HAWK, providing role based access controls for the remote multi-tiered administration over secure encrypted

sessions. The administrator can set roles based on multiple criteria ranging from group allowances to granular user roles that allows for the separation of confidential host information, and the destruction of system settings or otherwise sensitive data.

The HAWK Information Event Konsole, as the second piece of this solution, allows for historical retrieval of logged information up to a server side configurable period of time. This data is presented to the user in a logical fashion from highest priority alerts to lower priority alerts, all arranged by severity of correlation. This will identify to the analyst, exactly the source and trend of either attacker, or network problem.

Beyond this capability, the HAWK Information Event Konsole also has core functionality, which enables the capability to tune reporting for executive, remediation, and technical reports.

The HAWK Event Correlation Engine requires several separate components working in rhythm, accomplishing many tasks asynchronously. Each component has been assigned specific tasks amongst this complex multi-functional system.

1. The foundational component of the HAWK Event Correlation Engine is its relational database. HAWK comes packaged with both PostgreSQL and MySQL relational database support. System Architect Engineers with HAWK Network Defense, Inc. will help provide you with the on-site installation or training necessary for getting the HAWK Event Correlation Engine tuned and functioning properly.
2. The HAWK Event Correlation Engine (**HAWK ECE**) is necessary for collecting, filtering and correlating a myriad of sources providing from the most basic of logs to the advanced multi-relational database sources such as intrusion detection and prevention systems and single sign-on solutions. Each HAWK Event Correlation Engine supports failover with the help of a software or hardware load-balancer and must be logically located within the target network topology in order to properly identify both internal, as well as external internet addresses. The HAWK Event Correlation Engine initiates a securely encrypted session with HAWK's back-end database. This encrypted session allows the control of secure connections without the security risks of a remote virtual network connection.
3. The HAWK Information Event Konsole (**IEK**) is necessary for displaying and reporting the correlated event logs from the advanced multi-relational foundation. Each HAWK Information Event Konsole supports failover with the help of a software or hardware load-balancer and must be able to communicate with the HAWK Event Correlation Database.

HAWK utilizes the Naïve-Bayesian Histogram Analysis algorithm to uniquely “fingerprint” known security and performance issues, while establishing a baseline for positive or neutrally acceptable network traffic utilizing standard deviation. HAWK's unique “onion-layer” approach provides a stable, innovative

platform for applying a multitude of analysis techniques combining them into a unique “overall” score.

The technology provides a ‘single pane of glass’ making sense of thousands of events from over 50 vendor lines including Cisco, Juniper. It can expose and investigate hidden security threats in real-time with customized event correlation sensors tuned to the network’s unique activity patterns.

Reporting includes scheduled reports based on the client’s needs, detailing activity analysis, average event occurrences, and incident response time-lines.

HAWK provides several pre-packaged solutions:

- An optimized, cost-effective Intrusion Detected solution.
 - Multi-tiered hardware and software acceleration and support.
 - Up-to-date and advanced signature rule-set management.
 - Implementation of your own IDS/IPS infrastructure.
 - Integration with existing IDS/IPS monitoring infrastructure.
- Enterprise ACL Group Policies
- Long-Term Event Storage
- Additional Supported Vendors Formats
 - Tipping Point Intrusion Detection Event Correlation.
 - Radware Intrusion Prevention Event Correlation.
 - McAfee Intrusion Prevention infrastructure
 - Dragon IDS integration, training and support.
 - IETF IDMEF (Intrusion Detection Message Exchange Format) and many more.

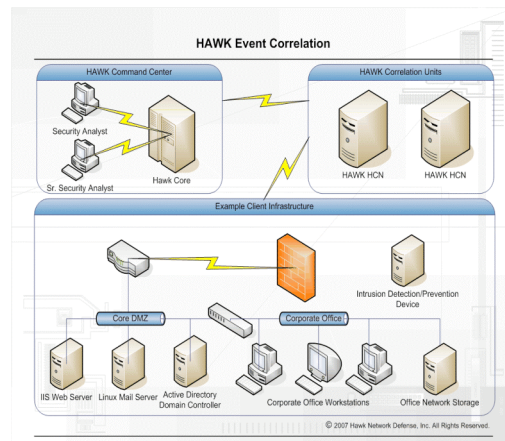
Event Correlation & Aggregation

An automated correlation tool collects reported event data from applications and infrastructure components, distilling them into a homogeneous set of events, which it in turn analyzes to reveal potential threats. Correlation is a background task performed in software, so it runs continuously for perpetual monitoring.

Advantages of Event Correlation:

- Decreasing response time for routine attacks and expediting root cause analysis. Traditional IT determination may take 3-12 hours; HAWK identification is less than 60 minutes.
- Potentially automating responses to reliably detected threats through automated e-mail notifications and/or SMS (cellular text message)
- Identifying suspicious activity without prior knowledge of attack.
- Gain a comprehensive business risk viewpoint of the network, quickly adapting to new and consistently changing threats.
- Streamlining data from multiple sources.
- Pinpointing bottlenecks and points of failure.
- Event Correlation and Real-Time Monitoring & Analysis
- Enterprise information assurance protection and management.
- Log/Event Correlation with support for over 50+ vendors.

- Patent-pending “Heuristic Learning & Trend Analysis”.
- Schedulable enterprise information report management.
- Multi-tiered enterprise “user/resource” access control.
- Provide your client the ability to protect its historical information between business groups.



Event Correlation Architecture

There are two basic types of correlation: rule-based and statistical. Experts suggest that combining rules and statistics in correlation maximizes effectiveness.

- **Rule-based correlation.** As its name implies, rule -based correlation depends on set definitions (rules) to relate events and analyze them in a broader context. A rule is essentially a scenario that an event must follow to be detected as an attack or a failure, applicable to both incoming events and to historical events stored in the database. A rules engine must hold events “in state” for a period of time until other qualifying events trigger an alert or the rule times out for the initial event.
- **Statistical correlation.** Statistical correlation relies on accumulated knowledge of normal events to identify patterns, which serve as points of comparison for new events. A pre-set algorithm calculates an incoming event’s threat level based on deviation from the historical norm.

The two basic types of correlation, either independently or in concert with one another, can be employed in a variety of methods, each with specific goals and focuses. Methods of correlation include:

- **Asset or target correlation.** This method places correlation in the business context by assessing events based on a prioritization of a company’s assets. To achieve this, various parameters—either user-defined or automatically computed from available event-context data—

skew the statistical algorithm, which then assigns priority to the most valuable business assets.

- **Comparative/Vulnerability correlation.** By comparing correlated or anomalous network threat activity to known baselines and prior vulnerability assessments, comparative correlation can track changes in threat exposure and risk level over time.

The Security Information and Log Management Markets

The Security Information Market

Security information management (SIM) products that contain event-log correlation tools streamline the threat identification and assessment processes by looking at individual events as well as sets of events bound by some common parameter. Research firm Datamonitor reports that the market reached \$174 million in 2004¹ and Forrester reports SIM tools are fast becoming must-haves for security teams wanting more visibility into IT activity within their environment. The market is currently growing at a rate of around 50%, and growth will continue to accelerate to \$1.13 billion by 2011.² Enterprises' needs to filter, aggregate, and correlate event information from multiple sources for real-time monitoring and historical analysis will fuel the projected growth.

Comparison of SIM Product Features								
Features	HAWKNet Forensics	GuardedNet	Open	ArcSight	IBM	Intellitactics	CA	
Aggregation	X	X	X	X	X	X	X	X
Rule Correlation	X	X	X	X	X	X	X	X
Standard Deviation	X	X	X	X	X	X	X	X
Bayesian Heuristics	X							
Non-Console Notification	X	X		X				
Threat Visualization	X	X	X	X			X	X
GeoIP Visualization	X			X				X
Reporting/Analysis	X	X	X	X	X	X	X	X
Performance Monitoring	X			X	X			X
Incident Response	*	X	X	X	X	X	X	
Policy Compliance	*	X	X		X			

Checks indicate advertised capabilities. Asterisk indicates capabilities in-development.

The Log-Management Market

Beyond SIM, vendors are including correlation tools in products designed to monitor network performance with increasing frequency. Known by a multitude of names—including log-management, performance-management, and performance-monitoring tools—these products aggregate event data from specific applications and/or infrastructure components, automatically correlating them to determine points of failure and performance bottlenecks. As the log-management market is an emerging niche market, it is difficult to pinpoint its value.

Comparison of Log Management Product Features							
Features	HAWK	Monolith	Consul Insight	OPNET	Empirix	NetIQ	Hewlett-Packard
Aggregation	X	X	X	X	X	X	X
Rule Correlation	X	X	X	X	X	X	X
Standard Deviation	X		X	X	X	X	X
Bayesian Heuristics	X						
Non-Console Notification	X				X		
Reporting/Analysis	X	X	X	X	X	X	X
Encrypted Historical Data Archiving	X						
Automated Response	*		X		X		X
Policy Management	*	X	X			X	X

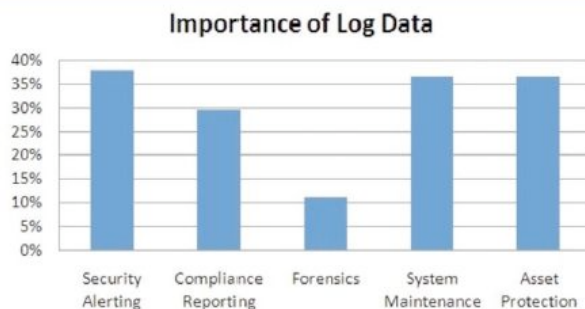
Checks indicate advertised capabilities. Asterisk indicates capabilities in-development.

Since nearly two-thirds of IT professionals are unhappy with their log data management systems, there is plenty of room for improvement. This unhappiness is mostly due to lack of correlation and normalization, areas the HAWK system addresses.

Even though more people are working with log servers and more of them are satisfied with their systems, two out of three are not deriving the information they need from their log management systems. It's clear that IT groups want to get more value from their log information. In a SANS survey, when asked how log data would most benefit their organization, respondents saw 'great benefit' for use of log data in event detection and tracking of suspicious behavior, day-to-day IT operations, process control/compliance, employee use monitoring, forensics, information leak protection and regulatory compliance. When broken down to Global 2000 respondents, *regulatory compliance* becomes a primary driver.

Survey respondents were asked to rank the three most important types of log related activities to their organization in order of first, second, and third choices. Leading their first choices was Information Asset Protection (46 percent), followed by system maintenance (35 percent), with security monitoring and compliance tying for 31 percent. Dominating their second choice was compliance reporting (36 percent).⁴

When you add up the number of respondents across their three choices (see following graphic) a slightly different picture is revealed. Looking at it this way, overall the data shows that respondents deem security alerting (38 percent), system maintenance and information asset protection (tied at 37 percent) and compliance reporting (30 percent) most important.



System Maintenance is ranked as a top use of log data. Based on the entire sampling, 62 percent say they collect log data to minimize downtime and assess IT incidents. In fact, the lack of log data is a serious obstacle to overcome when attempting to resolve system problems. End users often regard alerts and warnings as nuisances so they close those messages without recording the information. If the data is not logged someplace, it is necessary to re-create the error so that the support personnel can see exactly what happened. The availability of a full complement of log data from the application to the workstation, server and infrastructure can fill in the details of which the end-user may not even be aware.

Analysis of the Global 2000 reveals that the two top reasons for collecting data are archiving and compliance reporting, which are obviously related. Yet, based on their storage retention uses, organizations are not maintaining these logs indefinitely for compliance purposes. Most (14 percent) are unsure of how long they maintain their logs or they rely on the O/S default for that system. Just over 11 percent store their log information for 30-90 days, and a mere nine percent store their log data for six months or more. This is due to many factors, not the least of which is the sheer volume of data these systems produce and their lack of common format.

Compliance reporting is also a growing concern among respondents. In fact, our own research of 421 IT professionals put it exactly on par with security alerting and reporting. Over the past few years, regulatory bodies have considerably increased the requirements for logging of security-related data. Much of the data today required by regulations goes well beyond logs from their network and security devices. It also includes managing log data from applications where sensitive data might be stored and accessed by end users. This includes operating systems, databases, home grown and commercial applications, and mainframes. Tracking access to restricted data must become part of normal operation, as should the ability to tell when there is misuse of access to data. Survey respondents are collecting this data to varying degrees. Most (77 percent) are collecting firewall log data. After that, other forms of data collection drop off precipitously. Collection of antivirus, routers and IDS/IPS is done by 59 percent of respondents. At the application level, 58 percent are using their O/S logs, 57 percent are using their database logs, 46 percent use logs in their enterprise applications, 33 percent use logs on home grown applications, and mainframe application logs are used among 22 percent of respondents.⁴

In the SANS study, a majority (63 percent) of respondents are not satisfied with their current log file analysis processes. Over half (55 percent) of Global 2000 firms are unhappy with their log analysis processes, even though they spent an average of \$187,000 in 2006 on log management. Of our respondents, 89% would react favorably to an all-in-one platform to analysis, tracking and storage of log events, suggesting a product such as HAWK could have positive market potential.

Conclusion

Two key problems with log management are collecting data, and processing and reporting that data. These issues naturally stem from the sheer volume of log data, inconsistencies in log formatting and the lack of available logging features in many applications. Vendors of log management systems are working to resolve many of these issues, but in the meantime, IT staff members need to take it upon themselves to learn about their systems and the value of their log data

Ease of use, recognized as important to very important for 87% of respondents, is a key factor in an organization's use of such a tool, can be addressed through a logical graphical user interface. In the case of HAWK, a web-based frontend provides the administrator readily available access to all core features of the platform.⁴



HAWK Event Manager

Ease of installation, recognized as important to very important for 86% of respondents, is the second key factor in an organization's ability to deploy such a platform.⁴ The deployment of HAWK involves two to three components.

The HAWK Information Event Konsole (IEK) provides the secure web based interface for utilizing and managing HAWK. The HAWK IEK can run in parallel with other IEK's providing greater availability. By default, the relational database is stored locally on the HAWK Information Event Konsole (IEK) but can also be remotely located for enterprise implementations. Additionally, one or many HAWK's Event Correlation Engine (ECE) are required for gathering security and performance related information from its host network. The HAWK IEK and ECE are logically located on one device, allowing the end user to take advantage of both functionalities in a single platform, however the HAWK ECE can also be deployed in the similar manner to industry standards for Intrusion Detection/Prevention Systems. The HAWK ECE gathers information passively and is easy to configure for network operations. A simple console configuration wizard is provided for simple local configuration management, however most of

the HAWK Event Correlation Engine is pre-configured. The optimized tuning of HAWK is also available through modifying its configuration. This configuration provides an interface for tuning thread pools, data cache's and more. The TCP/IP Port 40001 is used for a secure connection between the HAWK ECE and the HAWK IEK database. Optionally, a large data storage solution such as a NAS/SAN will provide a solution for long term historical data archiving and regulatory policy compliance.

HAWK Defense, Inc. is a privately held corporation based in Dallas, Texas.
www.hawkdefense.com

-
- 1 Datamonitor, "Security information management: is it either software or managed security services?" January 6, 2005.
 - 2 Paul Stamp, Forrester Research, Security Information Management Market Forecast, 2007 To 2011,
 - 3 SANS 2007 Log Management Market Report, SANS Industry Analyst Team, Spring 2007
 - 4 Interminds, LLC, "SIM and Log-Management Survey 2008," 421 respondents, May 2008