Global Information Assurance Certification Paper

Security Information Management
Patrick J. Nolan Jr.
SANS GSEC Version 1.4b
September 1, 2003

Introduction:

Security Information Management (SIM) software uses a process in which dissimilar log file information is collected, normalized, aggregated, correlated and reported. The focus of this paper is to educate the reader regarding this type of software so that you can see why a SIM solution should be part of your security environment. This paper does not delve into the functionality of specific products or make comparisons, although in some instances specific products have been used as examples. Several product comparisons have been done and printed in such magazines as Network Computing[1] and Federal Computer Week[2]. Additionally, at the end of this paper is a sample of companies and their SIM software solutions. Further evaluation is necessary in order to decide which package will work best for your organization.

The geeks just want more toys to play with!

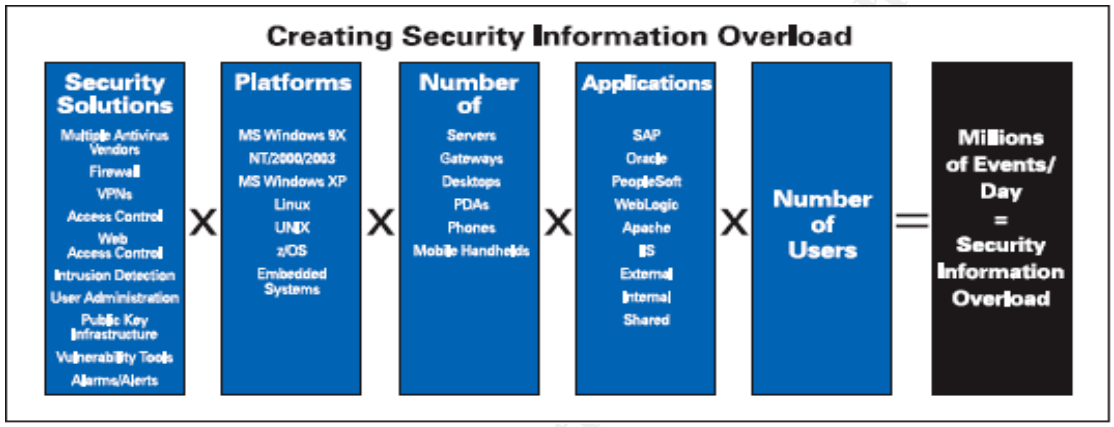
In the early days of computers, system administrators did not think about security. Their job was to make sure the systems were up and running. The priority was to be able to share files and all users were basically trusted. With the advent of e-mail, networks and rogue insiders, we started to have to defend against viruses. To solve that, anti-virus software was deployed so that administrators could monitor those logs for anything out of the ordinary. Then, as we started connecting external networks together, the techs insisted that a firewall would protect us from the outsiders. The next buzzword was IDS – Intrusion Detection Systems – host and network! Soon managing security became an out of control tornado, with torrents of disparate information that no one could keep up with. The idea of having all of these diverse, yet complimentary, layers of security technologies is referred to as defense-in-depth. Many experts have stated that a defense-in-depth approach to information security is the only way to properly and securely install and maintain our networks.

Defense-in-depth requires the deployment of different and overlapping technologies such as firewalls, IDS, anti-virus software, access control systems, etc. at different sections of the network. Each of these devices plays a vital role in defending our critical data. Their job is to see an incident occur, react to the incident and/or make note that something happened. This typically occurs by the device writing some sort of notification to a log file for the system or network administrator to review as they have time. But what if someone isn't available to sit and watch system logs? Due to the overwhelming workload required of system, network and security administrators it is not uncommon for this

information to never be reviewed. This situation is often referred to as information overload.

Information overload is a term commonly referred to when discussing information security. As an example of how a single event can generate thousands of messages see Figure 1. The only way to manage this successfully is to automate the process. This is where the maturing field of SIM technology can be of significant benefit to your organization.

Figure 1[3]



SIM solutions are designed to enable administrators to gather security relevant information from disparate sources into one central location in real time. By performing this function the security department is able to react to these incidents and to utilize the existing security resources more effectively. MANAGERS TAKE NOTE!! These resources are not just the networks and computer equipment that have been deployed but the highly paid security administrators that should be spending their time on other projects instead of manually reviewing log files for anomalies.

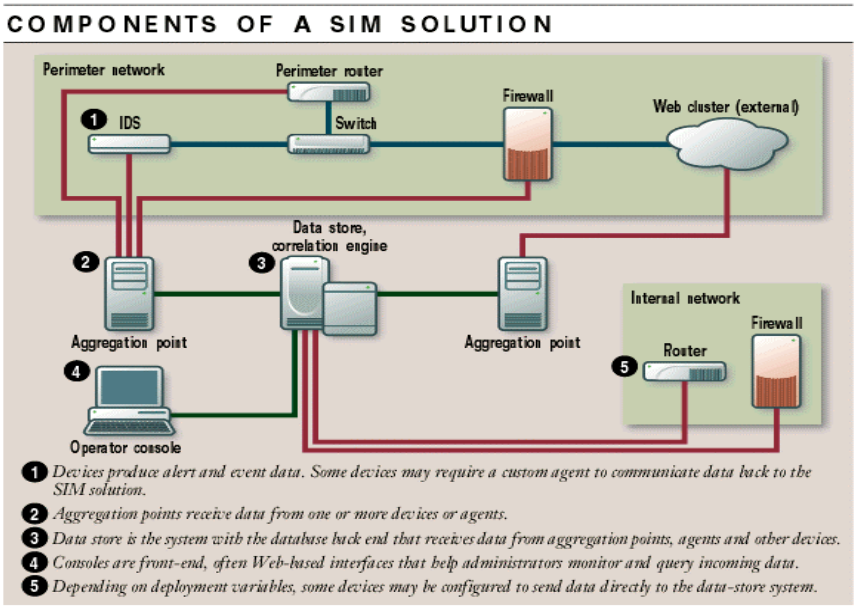
As you start looking at the cost of these packages you will find that it is not uncommon for the base software packages to cost in excess of $50,000. However, by comparing the cost of deploying SIM for your organization against what your personnel and other utility work is costing, you might be surprised to find that there can be significant savings in terms of real dollars and personnel productivity. Some of the SIM vendors have placed ROI (Return on Investment) calculators on their web sites for corporations to get a rough idea of whether SIM will benefit your organization in terms of dollars and cents[4].

Components of a SIM Solution

A SIM solution is designed to make the job of data consolidation and analysis a less arduous and more efficient task. There are three main components in any SIM solution (see Figure 2). First is the network agent, which focuses on normalizing and aggregating the log information. The second part of the solution

is the event correlation server, which tries to determine what action is happening and where. The last component is the reporting and logging mechanism. The information that has been previously processed is used to alert end users of events for manual or automated response. Additionally, this information is stored in databases and log files in several different formats for trend and forensic analysis. Now, I will break down these three components to explain what they do and why they do it.

Figure 2[1]



Agent:

As stated previously, the goal of the SIM agent is to normalize and gather system log file information in a format that can then be used to make informed decisions. For the purpose of this paper, we will focus on the fact that the information gathered is typically security relevant data. However, a decent SIM solution will allow implementation of many types of computer or network log information and not specifically security information.

Using a variety of protocols, such as syslog, SNMP, TCP/IP and vendor specific interfaces such as Checkpoint's OPSEC, the agent enables the device to communicate back to the correlation server. It is important to note that the type of protocol used may influence the ability of the agent to encrypt traffic between the agent and the event correlation server. "If the product relies on SNMP and syslog (as many do), then it is difficult to encrypt the traffic. SNMP is not a connection-based protocol so not only is encryption a problem, it is likely messages will be lost in heavy traffic situations"[5]. The reason that these two protocol examples may lose data during communication is because they use the User Datagram Protocol (UDP). UDP, as opposed to TCP (Transmission Control

Protocol), is not connection-based, which means that it sends data without any guarantee of arrival.

As you deploy a SIM solution, an important area of consideration is that of performance. Depending on whether you are already collecting log information, there should be little if any degradation of performance caused by the agents. However, if you are not already doing this you should investigate whether a hardware or software upgrade may be necessary. Depending on the agent, and the type of device that the agent is meant to monitor, this software component may be placed directly on the device that you want to monitor. This is not always the case though. For example, the netForensics agent for Cisco products resides on a syslog server.

The agents that are deployed should be able to communicate with any type of equipment or software that is able to gather security relevant data. Examples of agent technology are listed in Figure 3 provided by e-Security Inc. This is not an all-inclusive list and is only a representation of what some vendors may already have in place.

Figure 3[6]

Enterprise Resource Planning
- PeopleSoft
- SAP

Intrusion Detection (host-based)
- COPS
- Network Associates CyberCop
- Entercept
- Symantec Intruder Alert Manager
- ISS NetworkIce

Network Management
- BMC Patrol EM
- HP OpenView
- Micromuse

LAN Equipment
- Cabletron Switches
- Cisco Routers (all)

Operating Systems
- Windows NT/2000
- Solaris
- SunOS
- HP-Unix
- Linux

Intrusion Detection (network-based)
- Cisco Director
- Cisco Secure IDS (NetRanger)

Mainframe
- ACF2
- RACF
- Tandem
- Top Secret

Firewalls
- Checkpoint Firewall-1
- Cisco PIX
- CyberGuard
- IPChains
- Lucent Managed Firewall
- NetScreen
- Network Associates Gauntlet
- Sonic Wall
- Stonewall
- Symantec Raptor
- WatchGuard

Honeypot
- Symantec ManTrap

Integrity Assurance
- Okena Stormwatch
- Tripwire

Anti-Virus
- McAfee Virus Scan
- Symantec (Norton) Anti-Virus
- Trend Micro Mail Scan

- Enterasys Dragon
- Intrusion.com
- ISS RealSecure
- ISS SiteProtector
- Network Associates CyberCop
- NetScreen
- NFR
- Snort
- Symantec ManHunt
- Symantec Net Prowler
- Trend Micro Server Protect
- Trend Micro Virus Wall

Authentication
- Cisco TACACS+
- Radius Dial-Up Authentication

Databases
- Informix
- MS-SQL Server
- MySQL
- Oracle
- Sybase

Since there are so many different types of devices in place throughout networks it is impossible for SIM vendors to create agents for all of these devices. Fortunately, for every vendor researched, there is a mechanism in place that allows you to create your own agents. To help with the creation of these customized agents, most of the vendors have created agent wizards.

Security devices do not typically write log file information in the same format. In many cases they do not even call the events the same thing. The process of rearranging and renaming log file information to the same type of format is called normalization. The agent has to be able to normalize the information produced so that the security related data is sent back to the event correlation server in the same format that other security devices report back to the server.

For example, the following logs are from different network devices, which report on the exact same packet traveling across the network. These logs represent a remote printer buffer overflow that connects to IIS servers over port 80[7]. The fact that this example is a buffer overflow is irrelevant. The example is only meant to show how the different devices log an event. As you can see these logs have completely different formats. Without normalizing them first they would be of little value to an analyst.

Figure 4[7]

Check Point:
"14" "21Dec2001" "12:10:29" "eth-s1p4c0" "ip.of.firewall" "log" "accept" "www-http" "65.65.65.65" "10.10.10.10" "tcp" "4" "1355" ""
"" "" "" "" "" "" "" "" "firewall" " len 68"

Cisco Router:
Dec 21 12:10:27: %SEC-6-IPACCESSLOGP: list 102 permitted tcp 65.65.65.65(1355) -> 10.10.10.10(80), 1 packet

Cisco PIX:
Dec 21 2001 12:10:28: %PIX-6-302001: Built inbound TCP connection 125891

for faddr 65.65.65.65/1355 gaddr 10.10.10.10/80 laddr 10.0.111.22/80

Snort:
[**] [1:971:1] WEB-IIS ISAPI .printer access [**]
[Classification: Attempted Information Leak] [Priority: 3]
12/21-12:10:29.100000 65.65.65.65:1355 -> 10.10.10.10:80
TCP TTL:63 TOS:0x0 ID:5752 IpLen:20 DgmLen:1234 DF
AP Seq: 0xB13810DC Ack: 0xC5D2E066 Win: 0x7D78 TcpLen: 32
TCP Options (3) => NOP NOP TS: 493412860 0
[Xref => http://cve.mitre.org/cgi-bin/cvename.cgi?name=CAN-2001-0241]
[Xref => http://www.whitehats.com/info/IDS533]

So how could these events possibly be stored in a database productively? First, it must be decided which fields are relevant. Then you can develop a plan to accommodate the different fields that are populated by these devices. Choosing the fields must be content driven and not based on semantic differences between what Check Point may call a "target address" and what Cisco calls a "destination address." To accomplish the process of normalization, the agent is written to pull out those values from the event and populate the corresponding fields in the database. Figure 5 provides an example of a database containing these alerts after they have been normalized.

Figure 5[7]

Date	Time	Event_Name	Src_IP	Src_Port	Tgt_IP	Tgt_Port	Device_Type	Additional_data
21-Dec-01	12:10:29	accept	65.65.65.65	1355	10.10.10.10	80	CheckPoint	
21-Dec-01	12:10:27	list 102 permitted tcp	65.65.65.65	1355	10.10.10.10	80	Cisco Router	
21-Dec-01	12:10:28	Built inbound TCP connection	65.65.65.65	1355	10.10.10.10	80	Cisco PIX	
21-Dec-01	12:10:29	WEB-IIS ISAPI .printer access	65.65.65.65	1355	10.10.10.10	80	Snort	TCP TTL:63 TOS:0x0 ID:5752 IpLen:20 DgmLen:1234 DF ***AP*** Seq: 0xB13810DC Ack: 0xC5D2E066 Win: 0x7D78 TcpLen: 32 TCP Options (3) => NOP NOP TS: 493412860 0

These are the same four events described in Figure 4, except they have been normalized. This information can now be used to investigate an incident. With the data organized more efficiently, an analyst can pull all records containing a value that is of interest or sort by any field that may be relevant.

Now that we see what the vendors mean by normalization, it is necessary to reduce the amount of traffic that is sent across the network back to the correlation server. Without doing this, you could potentially saturate your network bandwidth causing your own denial of service. This information must be simplified and reduced so that common event messages are sent to the next phase of the SIM process, called aggregation. Event aggregation is the process of reducing the large volumes of event data into smaller, more manageable sets of information.

For example, a port scan against 6,000 firewall ports will produce 6,000 firewall event messages. These 6,000 messages can be aggregated into one message, which says that there were 6,000 of these events. This one message is then sent back to the event correlation server instead of thousands of messages. The ability to aggregate these messages is critical to reducing the overload that can occur at the event correlation server and across the network[8].

Event Correlation Server:

We have seen that the agent performs many extremely important tasks related to normalization and aggregation. However, the SIM "workhorse" is the Event Correlation (EC) server. What makes the EC so important and powerful is its ability to determine whether an attack is occurring by comparing all of the events that are occurring throughout the network in real time. The engine is able to determine this even though the information presented may seem to indicate a few benign and seemingly unrelated events. Depending on how the SIM solution is developed, there are three ways in which correlation occurs: rules-based, statistical based or a combination of the two.

This is another area in which network uniqueness plays a large part. The size and type of network that your SIM is deployed upon (i.e., LAN vs. Internet) will determine the number of EC's required to keep up with the workload. EC numbers and sizes are determined by calculating the number of events per second that the EC may need to process during a security event. As with the agents, it is important to size the solution properly so that critical event information does not get lost due to improper solution sizing.

Rule Based Correlation:

Rules based correlation is a predefined set of criteria, or rules, which enable the SIM to monitor for specific events, such as worms, viruses, buffer overflows and Distributed Denial Of Service (DDOS). The rules are used in conjunction with a rating system, which ranks the events in order of importance and alerts the system administrator when an incident occurs.

Figure 6 is an example of how we can create a rule using simple "if, then, else" statements to capture information regarding a *Domain Name Server (DNS)* attack. The DNS attack rule could be written as: **if** we receive a reconnaissance attempt from a firewall against the DNS server (DNS version checking or other connection requests), and **if** we receive one or more exploit attempts from an IDS against the same DNS server, **then** send a notification to the operator[9].

Figure 6[9]

Vendors typically employ graphical wizards to help ease the creation of correlation rules. Most vendors also give system administrators access to powerful scripting languages to create customized rules. The following is an example of an actual rule, using a vendor's rule programming language. The rule is triggered by an attack progressing through a firewall and executing on a targeted system. This code then works with the correlation engine to alert end-users or conduct automated responses.

filter(e.evt match regex("ile") or e.evt match regex("Accept")) flow window(e.sip = w.sip, filter(e.st ="N" and e.evt match regex("Drop")), 300) flow trigger(10, 300, discriminator(e.sip))

NOTE: Since correlation rules are considered intellectual property the author of these rules has asked to not be identified.

Rules can also be developed to automatically respond to incidents, such as pushing rules out to firewalls to block ports or IP addresses. Although this function may seem to be a significant benefit, it is something that must be deployed with caution. This is because automated responses sometimes have the adverse effect of blocking good traffic or installing system patches that may break applications.

Statistical Correlation:

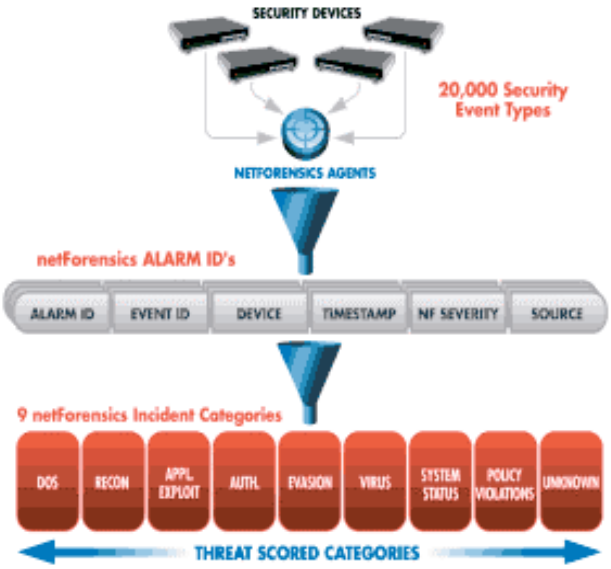
Thomas O'Laughlin at Symantec states,

> "Statistical Correlation relies on a representative, but valid sample size, to detect cyber attacks and predict how they will impact a particular environment."

Figure 7 shows how netForensics uses the normalized log file data and further processes it at the EC into different types of incidents. A threat score is then computed for each asset by combining event severity with the asset value that was determined for that system. Then, an overall measurement of security incident potential is determined. This method of correlation does not care what the specific attack is because it is looking at many events and using that information to form a potential attack scenario. The importance of anomalies that

may go undetected by only using a rules-based correlation mechanism cannot be understated, which is why the use of statistical correlation is so essential.

Figure 7[11]

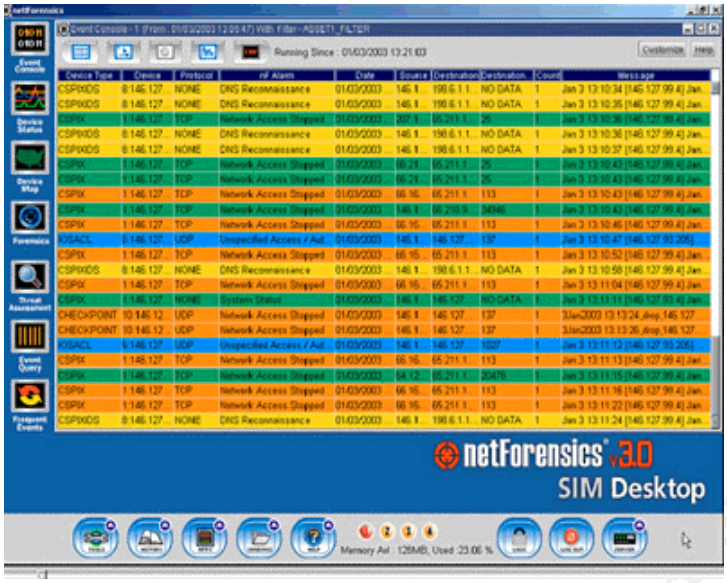


As previously identified, a limitation of rules-based correlation lies in its ability to detect only a finite set of security incident scenarios – namely, those that have been predefined. For this reason, an approach that employs both correlation methods in concert, enables the SIM to search for specific threat scenarios based on pre-defined rules, while making statistical measurements for alternative threat anomalies. This capability makes the use of SIM technology even stronger than just an event collection device. This ability, to correlate two or more unique, seemingly unrelated events, to determine whether an attack is occurring is the Holy Grail of Security Information Management[10].

Reporting and Logging:

Reporting and visualization are necessary to allow operators, analysts and managers alike, to gather the key information necessary to determine an overall threat posture, as well as investigate and respond to individual attacks before they affect or damage an enterprise. The master console allows for centralized detection and response to security events across the enterprise in real time. Notice in Figure 8 how easy it is to determine good traffic (green) as compared to events that are considered harmful (yellow and red). The job of monitoring this console for events does not require a highly skilled security professional. A less experienced operator can easily view events. Then, depending upon internally developed policies, an expert need only be alerted for the most important events.

Figure 8[12]

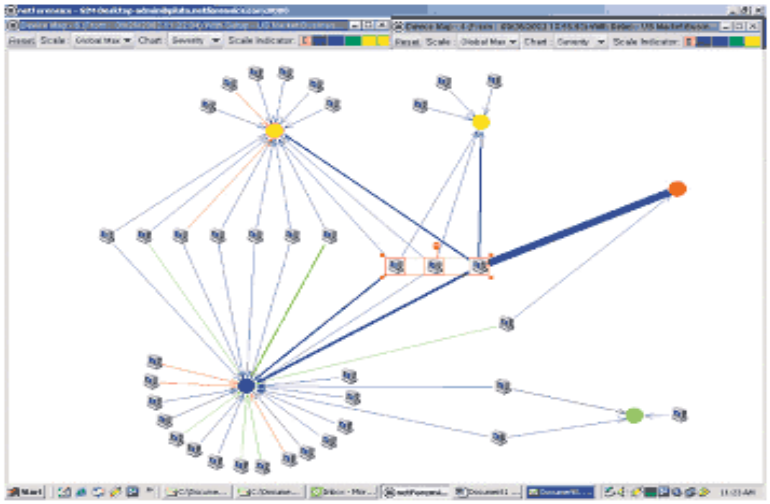


NetForensics Real-Time Event Console

Alerting is done via several methods. The event console is the primary monitoring station. However additional reporting methods such as remote consoles, pager notification, e-mail and cell phone messaging can be deployed.

Further investigation of these reports can be “drilled-down” through to further analyze the cause of the event, including the type of attack that was initiated, what the target was, if an asset was compromised and whether it was used to launch exploits against other systems or networks. All of this information can be analyzed down to whatever level (e.g., IP address, port, MAC address, etc.) is necessary to determine the offender, target and level of compromise. Due to the fact that the information is stored in a database, it can also be used for forensic analysis in the future.

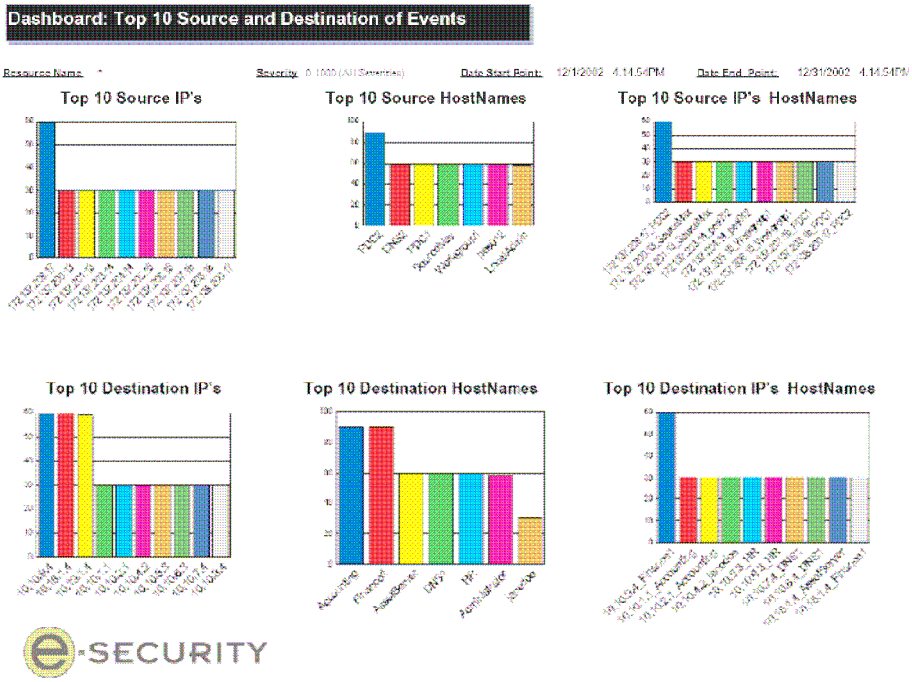
Figure 9[12]



On the previous page, Figure 9 shows an example of a different view that can be used to observe events. 2-D topographical maps such as the one shown, identifies correlated events, which are also color coded for ease of identification.

ROI (Return On Investment) is a topic that all CIO's deal with. How do they know whether the investment that they have made in a particular technology is helping or harming the bottom line of the business? By being able to provide management easy to understand reports, such as the examples in Figure 10, the security organization can show quantifiable results of what they are dealing with. The importance of this cannot be understated as managers are continuously looking for ways to improve efficiency and cost-effectiveness.

Figure 10[12]



Conclusion:

Information security is only becoming more complex every day. With attacks coming from multiple vectors and too much information from so many sources it is vital that an automated tool be employed to respond to events. Security professionals need some way to take back control of their networks. Security Information Management is a solution that you need to consider if you are responsible for the security of your organization's large and complex network.

Vendors and Products

This is by no means an exhaustive list of SIM vendors. The field is constantly attracting new players as software developers realize the need and potential for this important category of management tool.

NetForensics Inc.	netForensics®
Symantec Corp.	Symantec Incident Manager
Computer Associates Intl. Inc.	eTrust Security Command Center
Intellitactics	Network Security Manager™
ArcSight	ArcSight
GuardedNet	neuSECURE™
TriGeo	Contego™
e-Security Inc.	e-Security
OpenService Inc.	ThreatManager™

References

[1] Network Computing "Security Information Management Tools: NetForensics Leads a Weary Fleet" URL: http://www.networkcomputing.com/1307/1307f2.html (April 1, 2002).

[2] Federal Computer Week "Network forensics tighten security" URL: http://www.fcw.com/fcw/articles/2003/0707/tec-review-07-07-03.asp (July 7, 2003).

[3] Computer Associates "eTrust Security Command Center Brochure" URL: http://www3.ca.com/Files/BrochuresAndDescriptions/eTrust_SCC_broch_19599.pdf (July 11, 2003)

[4] GuardedNet "Product Brief: Calculating the ROI of a neuSECURE Implementation" URL: http://www.guardednet.com/literature.html

[5] Lunetta, Larry, VP Marketing and Business Development, ArcSight, Inc. via e-mail (August 6, 2003)

[6] e-Security "Agent Technology: Supported Products" URL: http://www.esecurityinc.com/products/agent_technology_supp_prod.asp

[7] ArcSight: "Data Normalization - The Foundation of Correlation" URL: http://www.arcsight.com/WhitePapers/login.jsp

"Got Correlation? Not without Normalization" URL: http://www.arcsight.com/WhitePapers/servlet/WhitePapers?step=5&MIME=pdf&file=5

[8] netForensics "Aggregation" URL: http://www.netforensics.com/documents/pr_sim_sublinks.asp?id=2

[9] netForensics “Rule Based Correlation” URL: http://www.netforensics.com/documents/pr_comprehensive_sublinks.asp?id=3

[10] GuardedNet “White Paper: Event Correlation: Is it Security’s Holy Grail?” URL: http://www.guardednet.com/literature.html

[11] netForensics “Statistical Correlation” URL: http://www.netforensics.com/documents/pr_comprehensive_sublinks.asp?id=4

[12] e-Security “Enterprise Security Management” URL: http://www.esecurityinc.com/downloads/Corporate_Brochure.pdf