Global Information Assurance Certification Paper

Copyright SANS Institute
Author Retains Full Rights

Interested in learning more?

Check out the list of upcoming events offering
"Network Monitoring and Threat Detection In-Depth (Security 503)"
at http://www.giac.org/registration/gcia

Distilling Data in a SIM: A Strategy for the Analysis

of Events in the ArcSight ESM

GCIA Gold Certification

Author: James Voorhees, voorhees.james@gmail.com

Adviser: John Bambenek

Accepted: 26 September 2007

Table of Contents

1 Introduction

The ArcSight Enterprise Security Manager (ArcSight ESM, hereafter, simply 'ArcSight') collects and normalizes network data. It can include data from intrusion detection or protection systems (IDS/IPS), firewalls, servers, web servers, and other kinds of devices, including routers and switches. The data can comprise millions of events. This dataset must be reduced so that analysts can make sense of it and find the events of interest that indicate that action must be taken. This is no simple task. Nor can it be done in a day. It must be planned, then carried out with painstaking care. There is, however, no guide readily available that will tell you how to do this.

This paper will give you a strategy that can be used to make that reduction. Its focus will be on how data appears to the security analysts whose job it is to look at the ArcSight console to find events that threaten their enterprise or violate its policies. It will not deal extensively with either the placement of sensors or with installation of the components of the ArcSight ESM. The paper assumes that they are installed and working properly, with feeds going into ArcSight. Nor will it be concerned with the preparation of data for business management or ArcSight administration.

2 Description of the ArcSight ESM

Anyone who has worked with ArcSight knows that it is big and complex. The documentation ArcSight provides runs to several hundred pages, yet still feels incomplete. What follows is an

Voorhees   3

outline of ArcSight's architecture, how it processes events, and the tools available for analysis. These will provide a framework for the filtering strategy described later.

2.1 Architecture

ArcSight can be thought of as having three sets of components, each working at a different layer (see Figure 1).

Figure 1: Architecture of the ArcSight ESM
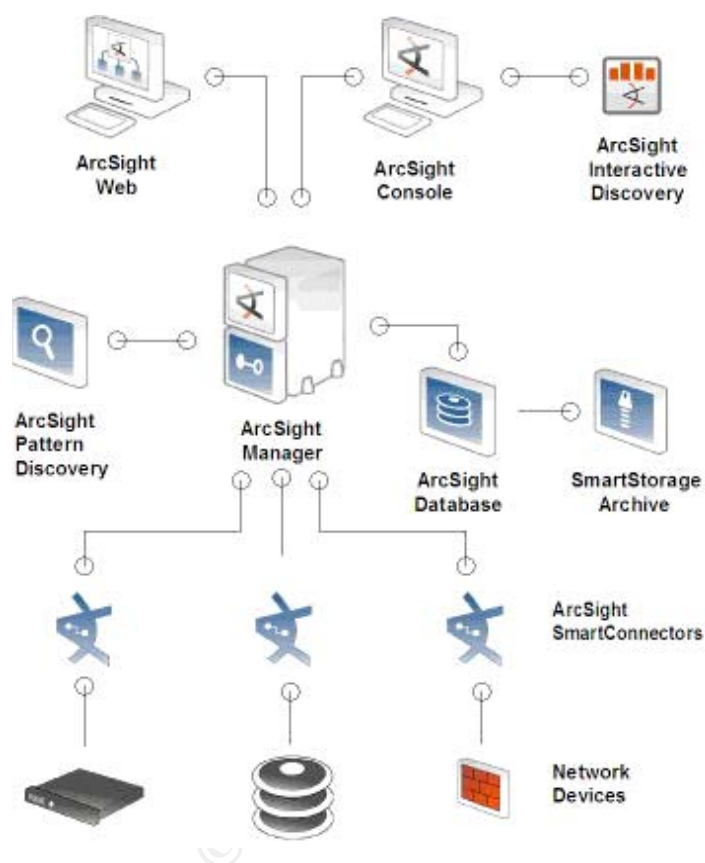
Source: ArcSight (2006, September 22).

Connectors: SmartConnectors (formerly known as SmartAgents; hereafter referred to as 'connectors') link ArcSight to devices throughout the network. The connectors collect events from the

Voorhees   4

devices and send them to a Manager.

Virtually any device on a network that generates data can produce events gathered by a connector.[1] That includes routers, switches, firewalls, servers, and host and network intrusion detection systems (HIDS and NIDS). The list of devices that ArcSight supports is extensive. If ArcSight has not created a connector for a device, a custom connector – a FlexConnector — can be created.

A connector can be installed on the device itself, on a separate machine dedicated to one or more SmartConnectors, or on a Manager. If a device sends information, the connector receives it. If not, it can be configured to retrieve the information.

The Manager: A Manager collects events from connectors. It then responds to requests, sending events the Console, which sits on a computer desktop. It also sends events to a database, which stores them. An ArcSight installation can contain a single Manager or several. Managers can report to other Managers, creating a hierarchy. Or they can be installed separately to cover separate functions or geographic areas. Each Manager requires its own database.

This paper will assume that there is a single Manager into which all events from all connectors are fed. If you have installed more than one Manager, the essential process will be

---

[1] For details on SmartConnectors, see ArcSight (2006, September 22). Also see Aquilar (2005).

the same, but the coverage and purpose of each Manager may differ.

The Database: The database is conceptually at the same layer as the Manager. The ArcSight database, which receives events from the Manager, is based on Oracle.

ArcSight stores data in three ways, as hot, warm, and cold data. Hot data is live data. Warm data is online but compressed into partitions of events collected over a 24-hour period. It is usually kept for between 30 and 90 days, depending on the policies of the organization. Cold data is offline, archived data. Cold data partitions can be stored on several kinds of media. How much of each kind of data you keep depends on how quickly you need to access it and the amount of storage space you have available.
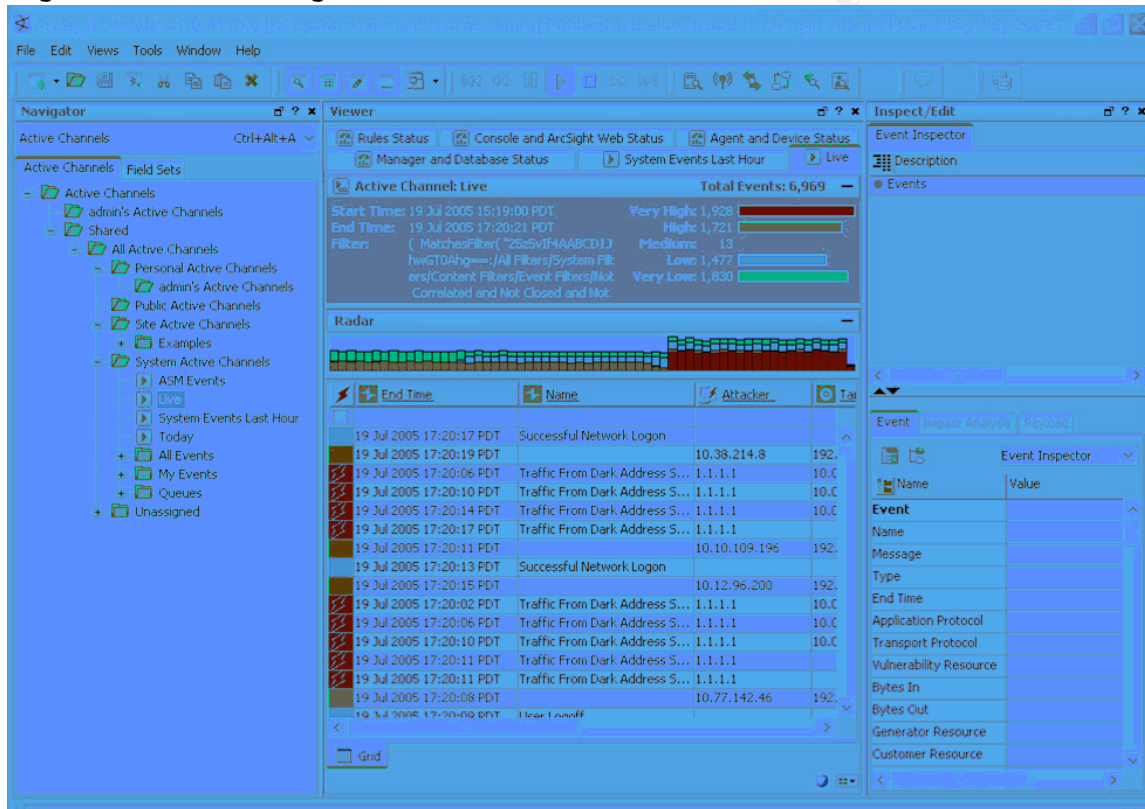
The Console: The Console and ArcSight Web give a user access to ArcSight, the latter through a browser. The Console has three parts.[2] From left to right as you look at it on your desktop, they are the Navigator, the Viewer, and Inspect/Edit (see Figure 2). Navigator is used to access what ArcSight calls "resources." Several resources are essential to tuning: Active Channels, Filters, Assets, Agents, and Rules. Other resources in the Navigator include Cases, the Knowledge Base, Patterns, and Users.

---

[2] For details on the ArcSight Console, see ArcSight (2005, November 29) and ArcSight (2006, February 9).

Events are seen in the central part of the Console, the Viewer. This is where Active Channels, Data Monitors, and other elements of what a user sees—his or her View—are displayed.

The Inspect/Edit panels, which appear on the right, are where changes can be made to individual resources. It is also where all of the fields in an event can be seen, even those that do not appear in an analyst's Active Channels.

Figure 2: The ArcSight Console

Source: ArcSight (2006, February 9).

ArcSight Web gives the user remote access to ArcSight. It lacks some of the functionality available through the Console, however. Because of this, this paper will focus on the Console.

There are other applications that can be added to the ArcSight ESM, as Figure 1 shows. Two of them add analytical capability, Pattern Discovery and Interactive Discovery. These build on the tools available in the console. Another, SmartStorage Archive, creates and organizes partitions. These applications do not help in tuning and so will not be dealt with here.

Turbo Mode: SmartConnectors and the Manager can be placed in one of three Turbo Modes. "Complete" mode provides everything that a device sends to a SmartConnector. The other two modes limit the amount of information that comes into ArcSight. ArcSight recommends using "Fastest" mode for "simpler devices such as firewalls." (Arcsight, 2006, September 22) "Faster" mode is the default for the Manager. The two advantages of the Faster and Fastest modes are that they require less storage space and greater throughput. The disadvantage, of course, is that the analyst will have less information available to determine what is happening on the network.

The Manager and the SmartConnectors do not have to be in the same Turbo Mode. The former will discard excess information sent by the latter. For example, a Manager set in its default mode — Faster — will throw out the extra information sent by a connector set in Complete mode. On the other hand, if the connector sends less information than the Manager is configured to receive, the Manager simply leaves the extra fields blank.

2.2 Processing Events

With that architecture in mind, how does ArcSight process

events? This process will provide a framework we can use to distill the events down to what an analyst can reasonably deal with.

Connectors: The process begins with the connectors. Before passing events along to the Manager, connectors aggregate some events and filter out others. These settings are set on the connectors themselves during installation. They can be changed from the Console (by those with the proper permissions).

The connectors also normalize the data that they collect. This means that they map fields set on the different devices to fields set in ArcSight. For example, different vendors use different numbers of fields and different formats to give the date and time. Checkpoint might use "24Jul2007" "12:10:29" to show what a Cisco router gives as Jul 24 12:10:29: and Snort presents as 07/24:12:10:29. ArcSight will change all these to "24 Jul 2007 12:10:29 GMT."

Normalization also includes translating the severity scales used by the different devices into ArcSight's "Agent Severity" scale. By default, this has five levels: Very Low, Low, Medium, High, and Very High. Each level has a different color associated with it. This makes it easy to distinguish the severity of events seen in the Viewer pane. The severity scale is one of several criteria used by the Manager to determine an event's priority, which can be an important element in tuning.

Connectors can be used to aggregate events. There are two kinds of aggregation. Basic aggregation creates one alert for a set of events that are exactly the same except for the time they

Voorhees   9

come in. It can be enabled on the Content tab of the Default tab
in Inspect/Edits. The time interval in which events are
aggregated and the minimum number of events to be aggregated
(the Event Threshold) can also be set.

Field aggregation creates a single alert for a set of
events that match on fields specified by the user. Any number of
fields can be used; those fields not specified are ignored.
Aggregation is disabled by default. It can be enabled and
configured on the same tab as basic aggregation.

ArcSight connectors also assign each event to a set of
categories (that is, it assigns a category tuple) using six
fields derived from the fields included in the events collected
by the connectors. These categories are designed to group like
events from unlike devices, from two different IDSs for example,
say, from ISS and Cisco. These six fields — Object, Behavior,
Outcome, Technique, Device Group, and Significance — and the
sub-fields under them form ArcSight's taxonomy for events
(ArcSight 2001; Marty 2005).

ArcSight determines the categories that events are assigned
to, except for events from FlexConnectors. The creator of the
FlexConnector must make that determination. ArcSight's
categorization is updated each week through Agent Update Packs
(AUPs). User-developed categorization will overwrite the
ArcSight-developed categorization.

Lastly, the connectors associate the events with a
"Customer" or "Zone." The customer tag is optional, useful
mainly to a Managed Security Service Provider (MSSP) who

Voorhees   10

distinguishes the events coming from different customers. Zones
are contiguous IP address ranges that can represent a group with
a common function or geographical location, such as a wireless
LAN, a VPN subnet, or your organization's Boise office.

The Manager: Once events are processed by the connectors
and sent to the database and the Manager, they are available to
the analyst. They can still be filtered out at the Manager. If
filtered there, they can still be retrieved from the database.
If not filtered out, they are sent to the Console. The Manager
can also do some processing on its own.

Every event in ArcSight is assigned a priority between 0
and 10, using a set of five criteria, most of which depend on
whether the asset has been modeled (more on modeling in the
Prerequisites section below). Prioritization can be useful in
filtering and is configurable (Saurabh, 2006).

Severity — Agent Severity to be more precise — set by the
connector, was mentioned above. The additional criteria are
Model Confidence, Relevance, Severity (not to be confused with
Agent Severity), and Asset Criticality (Saurabh 2006). Model
confidence assigns a rating based on whether an asset that is
the target of an event (that is, it is not the source or
attacker) has been modeled and scanned for open ports and
vulnerabilities. Relevance depends on whether the asset has been
scanned or not and whether the target has the targeted port open
or is vulnerable to the attack shown by the event. Asset
criticality depends on whatever criticality was assigned when
the asset was modeled. Severity depends on what is in several
Active Lists that are populated by system rules.

Voorhees   11

Each criterion is used to produce its own number between 0 and 10; these numbers are entered into a formula that results in the Priority for the event. The priorities can be customized by changing the filters and Active Lists used by ArcSight for prioritization or by simply changing the formula.

The Manager includes the correlation engine, which evaluates the normalized events for conditions set in filters, rules, and data monitors. It uses these conditions to determine what relationship events may have with each other, how significant that relationship is, what priority the events should receive, and what action to take. The action can mean sending a notification, creating a case, adding the event to an Active List, or even executing a command on the manager or a Connector.

The Console: Once events have left the correlation engine, they go to the Console, where they can be viewed (in the Viewer) in Active Channels, Data Monitors, or Event Graphs. This is where the analyst can make the events that come in from the Manager work for him or her.

3 Prerequisites

That is a snapshot of how ArcSight is put together and how it processes events to bring them to the analyst. As with most complex systems, there are a number of things that should be done—some as a part of planning, some as a part of preparing the system—before it can be run. Not all are technical; all are important if your use of ArcSight is to be effective. And all

Voorhees

will help in tune the flow of events to the analyst.

Risk Analysis: It is important to do at least an informal risk analysis to determine where the threats you face are coming from. It makes a difference whether you are concerned primarily with an external threat rather than an internal one, whether you have a VPN connection that could be targeted, whether you have users connecting over a wireless LAN, or whether you have one or many operating systems extant in your organization.[3]

Policies: Your organization should have a clearly stated, up-to-date set of policies that lays out what is forbidden and what is permitted. The policies need to be taken seriously and enforced. The importance of having such policies — both formal and effective — can scarcely be exaggerated. Cisco goes so far as to say that it "does not recommend deploying any technology without an associated security policy" (Dubrawsky and Saville, 2004). Policies define what is authorized and what is not; they provide users with boundaries for their activities on the network.

Your policies can tell analysts much of what they should

---

[3] One good method for conducting one can be found in NIST Special Publication 800-30 (Stoneburner et al. 2002)

The NIST publication describes risk management by "system," a term that can mean almost anything. Consider your "system" to be the network covered by ArcSight. That is, after all, what you are interested in. The analysis will not be a trivial effort, but the payoff will be to your entire security effort in addition to your effectiveness with ArcSight.

Voorhees   13

look for. If, for example, the organization limits the use of
peer-to-peer applications — either forbidding them entirely or
limiting them to one or two approved applications — then
analysts need to look at peer-to-peer events. If there is no
limit, analysts may want to ignore those events.[4]

   Procedures: Procedures will not tell analysts what to look
at, but they will tell them what to do when they find an event
worthy of examination. ArcSight can be used to establish a
workflow that can show how to handle any event, from a clear
false positive to a major attack. It can send an event to a
particular analyst or type of analyst. It can be used to
escalate events to senior analysts, incident handlers, or device
managers (firewall administrators, for example) for further
action.

   The different types of analyst will need to have different
sets of events in their Views in ArcSight. A junior analyst, for
example, might be charged with events that have been seen
before. Such events are familiar and the actions required are
known. Events never before seen might be given to senior
analysts. They could be given the task of finding out what the
event actually means and determining how to handle it. Incident
handlers and others analyzing events in ArcSight will,
similarly, each have their own set of events. Each type of
analyst will have their own standardized View to show these

---

   [4] SANS offers extensive information about policies, including templates,
at http://www.sans.org/resources/policies/. Also see Barman (2002)

Voorhees   14

events.

Log Settings: The log settings on the devices reporting to connectors so that ArcSight gets only the events that need to be seen.[5] For example, is it necessary to see both successful and failed attempts to access files and objects on all Windows systems? This can generate a large number of events, probably too many to be useful.[6]

Network Documentation: The more complete the documentation, the easier it is to determine which events are worthy of examination and what should be done about them. The documentation should show where the devices are that produce the events seen in ArcSight. It should identify the devices on the network and include as much information about them as possible, including the function of the device (firewall, web server, router, and so forth), operating systems, applications, and, where possible, a point of contact. It should include a list of critical assets so that you can give proper priority to the events involving those assets and focus your monitoring effort. Such a list can be an important element in the creation of ArcSight rules.

---

[5] Kochmar et al. (1998) provides a useful procedure for setting up logging.

[6] Microsoft provides a list of recommended audit settings for Group Policy in Microsoft (2006), Appendix B. Also see the best practices recommended for auditing security events in Microsoft (2005).

Asset Modeling: If you have reasonably complete information about the network and the systems on it, you can model your assets.[7] That is, you can create descriptions of the systems on the network. Such descriptions can include hostnames, IP addresses, MAC addresses, operating systems, applications, role, or criticality. The results of vulnerability scans from tools like Retina, Foundstone, and Nessus can be imported regularly. This data can give your asset models information about vulnerabilities and open ports.

In addition to making prioritization more accurate, modeling assets can allow you to filter events based on the operating system of the assets, the ports they have open, their vulnerabilities, applications installed, and the network or zone they are in.

Developed Use Cases: 'Use case' can be a fairly technical term, drawn from software engineering, with books written about it. Yet it is, in its essence, a simple concept. In this case, it means the scenario or scenarios you have for how to use ArcSight. You need to decide what you want to use it for, who is going to use it, where the information in ArcSight is to come from, and what will be done with the information ArcSight provides. One ArcSight staff member wrote that "ArcSight is not difficult if you break down the problems to simple statements first. When you do need to create content [rather than accept

---

[7] For an example how to model assets, drawn from the experience of UNISYS, see Theravukattil (2006).

ArcSight's defaults — JV] it is easiest to do so against a published agenda…" (DeStefano, 2007). The method he proposed is simple and should be familiar to many people working in information technology: define requirements, set goals, build content.[8]

ArcSight 4.0 can helps with this. It includes five "foundations" built on generic use cases that can be used to build more specific content that can serve your needs. Each includes resources that provide "real-time monitoring capabilities for its area of focus, as well as after-the-fact analysis in the form of reports, trends, and trend reports" (Arcsight, 2007, March 17). The foundations include content for ArcSight administration and workflow, configuration monitoring, intrusion monitoring, and network monitoring. When building you own content, it should also be clear what management needs for its reports, and what incident handlers, forensic investigators, and others in the workflow might need when events are handed off to them.

Use cases will determine how you tune ArcSight and how you set up Active Channels and Dashboards to show the tuned events. ArcSight can provide different kinds of events for different users. To pick an obvious example, ArcSight administrators will be interested in ArcSight events; your intrusion detection analysts may not. Less obviously, there are numerous events that

---

[8] Martinez and Veach outline a useful, detailed procedure for Developing use cases with ArcSight.

Voorhees   17

will be interest to a network administrator that have little to do with security. In addition, if you use ArcSight to gain compliance with a standard like PCI, Sarbanes-Oxley, or HIPAA, you have to determine what compliance will require.

Good Relations with Network Operations: The cooperation of the network operations group is vital if ArcSight is to be used effectively as an intrusion detection tool. There are good reasons for keeping the groups working on network security and network operations separate. They need to work together nonetheless.

This is important for tuning ArcSight because the operations group, unless it and the security are one and the same, has access to the devices that feed events to ArcSight. They can fix misconfigurations, one of which can account for many thousands of events. They can tell the security group what causes a particular event on a box. They can provide and update network documentation. Last, but by no means least, they can act on an event. They can change firewall rules or router ACLs. They can disable switch ports, stopping access from an infected computer or a user who has violated your organization's policies. If the group responsible for ArcSight has access to some of the devices that feed into ArcSight, all the better from this perspective. But seek or keep the cooperation of the group that runs the rest.

Because this paper is focused on intrusion detection, we will assume that ArcSight is being tuned for security analysts searching for intrusions and policy violations on the network. The workflow will run through analysts with different levels of

Voorhees   18

experience. We will not be concerned with the misconfiguration
of devices or with monitoring the traffic on the network for
outages, bottlenecks, or other problems not related to security.
These are important problems that ArcSight can help with, but a
security analyst does not need to see such events. Having said
that, the basic principles set out here should be a useful guide
to tuning ArcSight for such other uses.

In my experience one rarely has all of these. Sometimes you
don't have any. The worse any of these are, the more difficult
it will be to tune ArcSight well. If you don't have them, you
can't afford to simply throw in the towel. In many cases, you
cannot afford to wait to have these things done. Limits on time,
money and, let's face it, the motivation of the groups that must
provide you with these things may prevent you from getting them
in reasonable amount of time. This will be truer in large
organizations than in small.

4 Tuning in General: Principles and Process

The prerequisites listed above give you the means to know
your network and to understand what you want to do with
ArcSight. They give you the foundation you need to begin tuning.
Tuning ArcSight differs from tuning an IDS or server logs
because of its correlation capabilities. Events from one device
that mean little on their own can mean everything when combined
with events from others. For example, you may see thousands of
firewall events showing that connections were established and
then closed. The temptation may be to filter them out. Yet if

Voorhees   19

your IDS discovers an intrusion, that data may be precisely what
you need to find out when it occurred. It may be necessary to
keep them available to the analyst.

4.1 Principles

The fact is that it is difficult to know what events are
important until after you have experience with the system. You
should be able to filter out some events based on your
experience, your knowledge of your network, and your use case
analysis. Nonetheless, there is a risk that you may filter out
events that could be important to an analysis of events or the
investigation of an incident, either in themselves, or as part
of a correlation. This leads to the first principle of tuning
ArcSight: **If in doubt, don't tune it out**.[9] This may not be the
most important of the principles we will give here, but an event
that is tuned out, at least at the device or connector, can
never be used for analysis. If you think it might be useful
enough to keep, keep it until you know otherwise.

The goal of tuning ArcSight is to present the analyst with

---

[9] RSA's advice on log management serves just as well for ArcSight:

"Predicting what will be useful or not in today's environment is very difficult to
impossible. Wrong decisions can negatively affect audits or investigations. It makes
more sense to collect all of the data, and then review it to determine what you don't
need versus never collecting it. A well designed log management system can scale to
capture, analyze and manage very large volumes of log data, letting you collect all of
the data and intelligently purge whatever is assessed as unnecessary later."
(RSA 2007).

Voorhees   20

the events that he or she needs to take action on.[10] Only those events, and no others. This leads to the second principle: **Present the analyst only with events that he or she can do something about**. That something may be further investigation because they are suspicious or simply unknown. It could be that those events should be put into a case and escalated to, say, the incident handlers. That does not mean that you should filter out everything else so that it can be seen. Many events can safely be ignored most of the time yet become essential when the analyst investigates an event of interest. Those events must remain on-call. That gets back to the first principle: you do not want to leave your analysts without the data they need to understand what happens on the network.

This second principle and the basic limitations that analysts face as human beings, which means that no analyst can handle all the events coming into a network of any size, mean that tuning is essential. But where? Marcus Ranum argues that it is best to tune an IDS in a SIM like ArcSight, rather than at the devices the events come from (Ranum, 2004). A primary advantage of doing so is that you can centrally manage tuning. This not only makes tuning easier, it also reduces the work

---

[10] Jack Whitsitt wrote that "Basically, tuning the correlation engine (ArcSight) should never be approached from an "I need to get rid of stuff"— pure data reduction—standpoint" (Whitsitt, 2007). Instead, to paraphrase him loosely, you should use the tuning process to understand the environment, that is, the what, when, why, and how of communications on the network. The reduction of data will be a beneficial side effect of this.

necessary to install a new device: you can simply accept the
default settings for the signatures or logs.

Another reason for tuning all data—not just IDS data—after
it reaches a SIM is that the data, when taken as a whole, may
add to what you about what is going on your network. What
Richard Bejtlich calls statistical data--data that summarizes
network traffic by category to show deviations from norms
(Bejtlich 2006)—can show problems with the entire network, a
part of the network, or one device on the network. A surge in
the number of events coming from a host, for example, may signal
an infection.

As the tuning progresses, the reasons for each decision
must be documented. This third principle is essential: **Document
what you do.** This will make it possible for those looking at the
configuration later to make sense of it. Rather than wonder
months later why a rule was made, why an IP address was placed
on a Hostile list, or why SNMP events from 3.3.3.3 were filtered
out, it will save time and energy if such things are documented
when they are added.

Many ArcSight resources, including filters, cases, and
rules have a Notes tab. This is a convenient place to document
what you do. The Knowledge Base is somewhat less convenient, but
can be a useful place for one or more documents showing what was
done.

A fourth principle may seem obvious: **Tuning is iterative**.
This means, for one, that the process of tuning will never end.
Each time a new device is added to the network or a signature

Voorhees   22

set is updated, tuning — fine tuning, if you will — needs to occur. The new device may bring new events into ArcSight. A new signature set certainly will. You will need to find out what has changed and work it into what has become, in effect, your baseline. You can also tune out later the events whose usefulness you had doubt about at first. That first principle above does not mean do not tune events out; it means, understand them before you do.

This fourth principle can also guide how you approach tuning. This can be especially important if the computer resources available to ArcSight — CPU cycles, memory, and, especially, database space — are limited. The essence of this approach, explained in more detail below, is to take the events piecemeal, systematically, broadening the time you allow for events to come in and the devices they come in from until you have the entire system operating as it was designed to. A systematic approach will make it easier to handle the task of tuning without being overwhelmed by the number and kinds of events coming in. As you learn about a portion of the set of all events, you add another.

4.2 Process

Those are general principles. There is also a general method that can be adopted.[11] It is time-consuming — how much

---

[11] This method is adapted from the tuning procedure for network IDSs outlined in Cisco Systems (2004).

time depends on the size and complexity of your network — and complex. But it is necessary.

Determine the meaning and cause of the events seen. This is the key part of the tuning effort. It is no simple task. Its speed and effectiveness depend in large measure on how much the people tuning know about network traffic. They will also depend on the quality of the network documentation.

Those looking at events will need to take account of their sources and destinations (or, as ArcSight shows it by default, the Attackers and Targets), and the destination ports.[12] The source and destination addresses will tell which interface the events come from or go to. The network documentation should tell whether the interface is inside or outside the network; if it is inside, it should tell which machine it is and where to find it, at least on the logical network. If you cannot get its physical location, the network operations staff ought to be able to. The destination port will suggest which service the event is about.

Other information can also be helpful. Bytes In or Bytes Out, for example, can indicate whether the event fired on a packet of unusual size. The events in ArcSight do not always

---

[12] In ArcSight, the Attacker and the Target, a product of normalization, are usually the same as the Source and Destination. Differences are not usually seen. Conceptually, however, the two are different. As ArcSight puts it, "The values in the source and destination fields characterize the flow of traffic on your network, but "That source and destination may become an *attacker* and *target* if a network analyzer, such as a HIDS or NIDS, evaluates the traffic as hostile."

Voorhees   24

reveal whether the protocol used is TCP or UDP, but of course it can make a difference which one it is. For example, it can make a difference whether DNS traffic (port 53) is a normal request or reply (UDP) or might be a zone transfer (TCP).

Vendors' documentation can be essential. Most IDS vendors have lists that give the names and meaning of events. Good network documentation will help as well by making it easy to find out what type of events should be coming from a device. Some IDSs will tell you what the events fire on. Intrushield, for example, looks for one of several strings to show that an executable is present in an email attachment. Unfortunately, the documentation a vendor gives will often be frustratingly incomplete. More than once have I found the reason for an event given as something like "protocol error" with no indication what the error was.

Moreover, whether the event is malicious or not also depends on context. I once saw an event that indicated that Metasploit had been placed on a system, only to find out that a user had used the word "Metasploit" in the text of an email. The signature had fired on that word. In several cases, the event from an IDS warned me that malware may have come in through an email attachment, only to find that the executable was completely benign, perhaps a zipped file given an ".exe" extension. Intrushield had fired on one of the strings mentioned above. Yet at other times that same IDS event did, indeed, point to malware. In another case, the IDS warned of a UDP port scan. But it fired on DNS replies, with a source port of 53 and multiple destination ports that happened to be low ports (less

Voorhees   25

than 1024), not the usual high ports.

Nor should you neglect Google. After all, "Google is your friend." It is an essential source for information on network events. It helps fill gaps left by vendor's documentation and the knowledge of the analyst. No one person, after all, can know everything needed about the myriad protocols and applications found on a typical network.

The network staff should be a good source of information about what events you should see. In fact, they may have done some of the work on the events appearing in the logs before ArcSight appeared. Regular examination of system logs would have yielded some knowledge of what might be useful and what not.

Determine whether devices can be reconfigured to stop the activity behind the event. Clearly this cannot be done with the establishment and closure of authorized connections between devices. But the events brought into ArcSight can reveal myriad misconfigurations on a network. Naturally, the cooperation of the network operations staff is essential here. If the cooperation with the network staff is less than wholehearted, the ability of ArcSight to detect misconfigurations can be used as a carrot to foster it.

Aggregate or filter events. Finally, with events understood and misconfiguration removed, you can begin to reduce the number of events, to whichever level you decide is best, using whatever method fits, whether aggregation or filtering.

Aggregation is done to similar events that should be kept. Like filtering, it reduces the number of events seen. It can be

Voorhees   26

done through connectors or rules. We will have more about it below.

Filtering is more complex and will take more time. A systematic approach is needed.

5 Getting Started With Filtering

For most ArcSight installations, filtering will be no small task. The time allotted to it will vary by the amount of events that come into the system, the number of person-hours available, and the experience of the people doing the tuning.

A key to filtering effectively is to avoid having the people tuning and the ArcSight system overwhelmed by what can be a daunting task. As with any large task such as this, it becomes more manageable if you make it smaller by dividing it into its constituent parts.

In tuning ArcSight, you can **limit your data by type of device.** Begin with, say, events from firewalls alone. Use the tuning process outlined above to filter out the events that you need not see. Then add another type of device, router logs, perhaps. Do the same thing. Continue the process until all devices have been entered into ArcSight.

A different approach would be to **limit data by location**. You could bring in all data from a particular subnet, for example. It would easier, however, to be able to identify events coming from one particular kind of device at a time. Each device and each vendor, after all, have their own format for events and their own names for events. Normalization notwithstanding, you

Voorhees

27

have to learn how to interpret each type of events.

If limiting data by device or location overloads your system, you could **limit your data by time**. That is, you could only accept events from a fraction of the time that you intend to allow for hot data, the events immediately available to the analyst. If you intended to make 30 days of data available to the analyst, accept only 10. Work through this set of data, filtering out what needs to go, then allow more data and go through the process again. Continue until your system operates as it was designed.[13]

6 Tuning the ArcSight Manager

Having chosen the sets of events to begin with and having completed all preliminary tasks, you can then turn to the lengthy task of separating out events. Tuning should begin with the Manager, using filters, rules, and Active Channels to go through the steps outlined above. The resources needed — the filters, rules, and Active Channels — can be seen and changed on the Console.

Rules and Active Channels are built on filters. Filters, therefore, are the basis for tuning. A good way to begun tuning

---

[13] Justin Wilder offers some sound advice: "Rules and filters can be created by grouping events from a known incident (detected through another system). This can be an excellent source of filtering/tuning information. This would be an example of tuning "in" events. Table top exercises or attack modeling can also serve as an excellent source of information" (Wilder, 2007)

Voorhees   28

is to create an Active Channel that shows only the events from
the devices or locations that provide the events you are tuning.
This Active Channel — let's call it the Tuning Channel — can be
the primary way to view the data while tuning.

Then create two sets of filters: one for false positives,
the other for misconfigurations. Create two folders in the
Navigator for these filters, one for each set. The URIs might be
/All Filters/Public Filters/False Positives and /All
Filters/Public Filters/Misconfigurations. Create a meta-filter
in each folder and add each new filter to it as you create it.

Meta-filters filter on all the other filters in their
folder. When seen in Inspect/Edit, the filter will contain
several instances of "MatchesFilter("/AllFilters/[FILTER URI]"),
where [FILTER URI] is the address in the Filter part of
Navigator of other filters in the meta-filters' folder.

Add the meta-filters to the Tuning Channel so you can see
how far tuning has taken you. The meta-filter in the False
Positives folder could be named _False Positives_. Two of the
filters in the folder might be "DNS: Domain Controller Noise"
and "Domain Controllers: Kerberos Events." The filter for _False
Positives_ would then include these two entries:

MatchesFilter("/AllFilters/Public Filters/DNS: Domain
Controller Noise")

MatchesFilter("/AllFilters/Public Filters/Domain
Controllers: Kerberos Events")

A naming convention should be used to distinguish the meta-
filters from the ordinary filters. There are numerous ways of

Voorhees   29

ding this. Underscores (_) or other punctuation or putting the entire name in capital letters are but two. Underscores will be used in the examples here.

The false positive filters will be for events that are known to be benign. You should create a separate folder within _False Positives_ for events of no further use in investigation. You can use this to distinguish the events that you can later filter out at the device or connector. An example might be events that show 'pings' against the external interfaces of a firewall. A separate filter for each Connector will make it easy to add the events to the filter on the Connector later.

The other events in the False Positives folder will be those that an analyst might need when an event warrants further investigation. Events that show that a firewall denied a connection, built one up, or tore one down are an example of the latter. They usually have little value in themselves, but can provide valuable history for the analysis of an intrusion.

False positives should include events that are a part of normal network activity. Events that might be filtered out include legitimate NTP events, DNS queries and responses, and broadcasts (Babcock, 2006). Cisco's SAFE documentation presents a precise guide for filtering by segment of a network (Dubrawsky and Saville, 2004). In general terms, following their advice, you should filter out events that are allowed. A properly written policy and good network documentation can help significantly. Much of the advice in Cisco's documentation depends on knowing precisely what traffic a device should see. This depends on the device's function and location in the

Voorhees

30

network, what is allowed, and what is normal.

ArcSight produces two kinds of events that are useful for administering the tool, but not for intrusion analysis. These are audit events, produced by the correlation engines when a rule or data monitor triggers an action, and status monitor events, called ASM events, which are generated by the Manager for reporting and troubleshooting. The filters 'ArcSight Audit Events' and 'ASM Events' found in 'All Filters/System Filters/ArcSight System Event Filters' can be used to filter these events out.

When filtering, do not make the mistake of filtering out all encrypted traffic. It is true that the NIDS, and ArcSight after it, will be unable to read the payload. But data about the session—source and destination addresses, ports, times the session was built up and torn down—can all be important both in detecting an intrusion and in investigating it later. A session with a suspicious host is itself suspicious. One to an unexpected port should be investigated. A session that lasts too long, that happen too frequently, or exchange too much data ought to be looked at.[14]

---

[14] Bejtlich (2006). Bejtlich calls this, not surprisingly, "session data." We mentioned "statistical data" above. He also refers to "full content data," which is essentially the complete packets crossing the network, and "alert data." The last are the events that we are concerned with in ArcSight. This is a useful taxonomy of the data that can be used in intrusion detection. Also see Bejtlich (2005).

Misconfigurations are what the name suggests—those events that can be removed if the configuration of a device can be corrected. It would be better to have network operations fix the problem, but the events can be filtered until they do.[15]

Once you have finished the first set of devices or locations, replace them in the filter to the Tuning Channel with the second set and repeat the process. You need not create new meta-filters. Nor do you need to change the meta-filters in the Tuning Channel. Continue in this fashion until you have worked on all devices or locations.

With that done, basic tuning will almost be complete. However, the set of events remaining will not satisfy the second principle. It still contains events that are not needed for analysis—that is, it contains events that the analyst, because of practice or policy, can do nothing about. It also contains events that are not needed for either analysis or investigation.

Three remaining steps will distill the events further, leaving ArcSight with only what the analyst can use. One is simple, tuning events out at the connectors and devices. The work done to this point should leave you well prepared for it. The second is tuning the Manager to minimize and highlight the events that the analyst finds of interest. Rules are an essential part of this. The last step, which we can only touch on, is up to the analyst, who must manipulate and filter events

---

[15] Network operations might disagree whether it is a misconfiguration, of course. It can then be moved elsewhere, such as a False Positives filter.

to make their meaning clear. ArcSight provides myriad tools for this.

7   Tuning Devices and Connectors

Your work on the Manager should have told you which events can be tuned out before they enter the ArcSight database. That is, events that can be tuned at the device or the Connector. These will be false positives that are likely to be of little use in any investigation of events of interest. Filter events out here after you have determined that they need not go into the database. The folder within the False Positives folder should contain all such events that you have found. The rest of the events, those not filtered out and those filtered out at the Manager are still retained in the database, available for the detailed analysis of events of interest.

You have a choice of filtering events out at either the Connector or the device. Filtering them out at the Connector can make it easier to manage the events, if the Connector takes its events from more than one source. If the Connector takes its events from a device that itself manages devices, there may be little gain from tuning at the Connector. Filtering events at the device can save you bandwidth by keeping the events from crossing the network between the device and the Connector.

Voorhees   33

Figure 3: Connector Filters
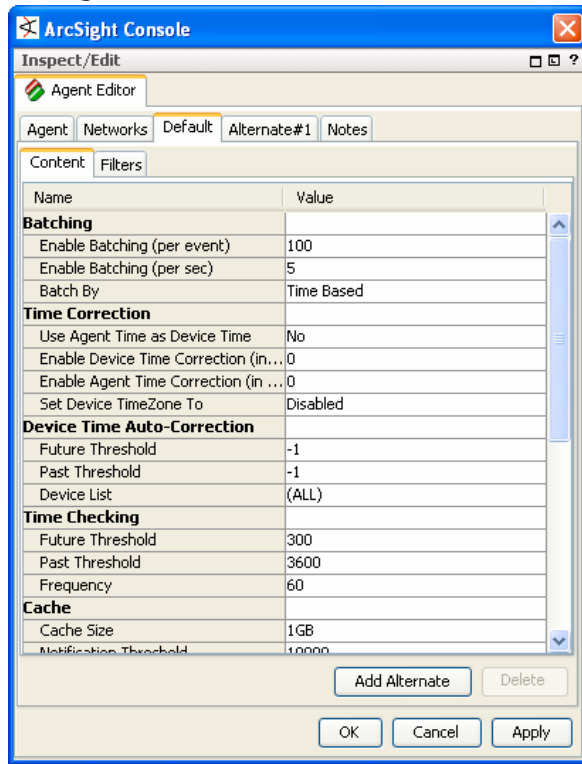
Source: ArcSight (2006, September 22).

The way events can be filtered at the device varies by
vendor, of course. To filter events on an ArcSight Connector,
bring up the Connector (or Agent) in Navigator. Go to the
Default tab in Inspect/Edit (see Figure 3). The Filters tab is
one of the two choices that then opens up (the other one is
'Content'). If you created a separate filter for each Connector
in the folder containing False Positives to be tuned at the
Connectors, you can add that here by selecting Filters and
browsing to the one you need. Otherwise, you can add conditions
to this filter as you would to any other.

After you have filtered out these events from the devices
and Connectors, the events remaining will be available to the

analyst. They can be viewed by opening an Active Channel with no filter.

8 Preparing Events for Analysts

After events are filtered at the Manager and filtered out at the Controllers and devices, the events flowing into the database should be tuned properly. The main task remaining is to tune the events seen by the analysts in their View. This View will include Active Channels primarily, but should also include other analytical tools such as Dashboards, Event Graphs, and Pattern Discovery. This may require changing the structure of filters from the simple one created for basic tuning to a more complex one to fit the analytical environment. It will require developing rules.

Rules correlate events and create new events to report on those correlations. Rules also aggregate the events that make the rule fire. This is significant in reducing the number of events for analysts. You set conditions in the rule. Conditions can be a filter, or an asset or vulnerability (if you have populated the tabs under Assets in Navigator), the fields and time for aggregation, and the actions the rule should take.[16] The actions can include creating cases and trouble tickets. Rules

---

[16] For writing rules, see ArcSight Help and ArcSight (2006, February 9). A good practice for creating rules and other resources in ArcSight is to use the content that ArcSight provides as templates. An essential practice is to test them thoroughly.

can also populate Active Lists and set event fields. The last
can be used in filters for the events that appear in an
analyst's View.[17]

The golden rule to building rules and any other content in
ArcSight rule is: Start with filters (Martinez, 2006). You
should create a structure of filters that can support the
categorization of events that you choose. This can be ArcSight's
event categorization scheme, described below. Or it can be the
categorizations of the Department of Defense or the Federal
Agency Incident Categories of US-CERT (US-CERT, 2006).

You should create folders for each set of filters that you
create, if they are not already a part of ArcSight's default
content. If you have several customers, you can create a set for
each. Folders can be created for each type of event. For example
the Department of Defense has seven categories for events. US-

---

[17] One way of populating Active Channels with events from rules is to
populate Category fields when setting Actions for the rules. One group used
the DeviceCategory field and populated it with a hierarchy of categories,
similar to an example like this:
/Security/SOC/IntrusionDetection/JuniorAnalysts

This allowed the group to create filters in Active Channels that would
fill the channel with, say, events designed for all Intrusion Detection
analysts or only junior analysts in the Intrusion Detection section.

Note that this group rarely used categorization. It also chose a field
that ArcSight's categorization did not populate with anything that the group
found useful.

Voorhees   36

CERT has a similar categorization, with six categories. A folder can be created for each type. Within the folder, you can create a meta-filter that can be used in a Rule or Active Channel, rather than each of the other filters in the folder. Among other things, this will make it easier to add or subtract filters.

The rules and filters can also be structured to reflect the Views that you want to create. Putting rules and filters into separate folders for, say, Tier 1 and Tier 2 analysts, or for Intrusion Detection and Incident Handling can make it easier to create meta-filters and point events to the proper Active Channel or Data Monitor.[18]

ArcSight's set of categories for events has a broader purpose than the two taxonomies noted above. ArcSight tries to place every event that appears into a common taxonomy so that they can be used generate for analysis and reporting. As noted above, each event is placed in a set of categories that has six dimensions: Object, Behavior, Outcome, Technique, DeviceGroup, and Significance. The categories for an event can be seen in the Categories section in the Event Inspector tab in the Review/Edit panel after you right click on the event and hit Show Event Detail. The following categorization, for an attempted brute-

---

[18] A folder that contains filters that can bring out undetermined events can be helpful, particularly if you adopt a tiered system of analysts, with senior analysts tasked with determining what such undetermined events are.

Voorhees   37

force login to an operating system, provides an example:[19]

Object:
/Host/Operating System
Behavior:
/Authenticate/Verify
Technique:
/Brute Force /Login
DeviceGroup:
/IDS/Network
Outcome:
/Attempt
Significance:
/Compromise

These categories give you a basis for filtering. You can include or exclude on the basis of what appears in that Category section. A filter on /Host/Infection/Worm, for example, would bring up all events that ArcSight had labeled such. A filter on /Host/Infection would bring up all events that ArcSight had labeled as infections.

With rules and filters structured as you would like them, you can create the Active Channels and Data Monitors that will present the analysts with the events that they can and should do something about. There are many ways of doing this. How you approach it depends on how you have defined your use cases, which will have defined who needs to see what.

---

[19] The example comes from Marty (2005). For details on Event Categorization, see that and ArcSight (2005, May 13).

9 Tuning for Analysis

Once ArcSight is tuned at the devices, the Connectors, and the Manager and the Active Channels and Dashboards for the analysts have been created, analysts will be presented with the events that they need to deal with. Importantly, the analyst can see all events coming into the Manager simply by opening up an Active Channel without a filter. That is why it pays to be careful when filtering events out at the device or connector, before they come to the Manager.

The analyst has numerous tools at hand within ArcSight that can be used for the further investigation of events. These include Filters, Rules, and Data Monitors, but also Pattern Discovery and Interactive Discovery. How to use them is beyond the scope of this paper.[20] Over time, as the analysts gain experience and training and work with each other, they should develop standard procedures or at least a set of best practices for dealing with different kinds of events.

10 Conclusion

The complexities of ArcSight can make it seem overwhelming. Yet by taking the time to define clearly how your organization will use it and then adopting a systematic, small-chunk approach to tuning will help ensure that analysts, and others, can use

---

[20] For a useful guide to developing ArcSight content for analysis and reporting, see Thomas (2007).

ArcSight to its full potential.

Tyson Foods spent several weeks getting devices to send data into ArcSight (SANS, 2007). Realistically, tuning should take weeks as well. How many weeks will depend on the abilities of the analysts and the quality of the preparation. As often happens with information technology, when time is taken to plan and prepare things well up front, it becomes easier to perform well later.

11 References

Much of this paper springs from the experience of Justin Wilder and Jack Whitsitt. My thanks to Jack for a thought-provoking, insightful email message on the paper and to Justin for his invaluable answers to numerous questions and for detailed comments on the text. Thanks, too, to Frank Olmstead for his invaluable comments and corrections. Any errors that you might find, naturally, are mine alone.

Aquilar, Hector (2005). *Introduction to Agent Architecture*. ArcSight 2005 Users' Conference, North Bethesda, Maryland.

ArcSight (2002). *Planning and Executing Enterprise Security Management*. ArcSight White Paper. Retrieved 9 September 2007 from http://www.genesiscom.ch/de/2/products/downloads/Planning and Executing.pdf.

ArcSight (2005, May 13). *ArcSight Event Categorization: A Technical Perspective*. ArcSight ESM Version 3.1, Revision 1.1. Cupertino, CA: ArcSight. Inc.

Voorhees   40

ArcSight (2005, June 28). *ArcSight Use Cases: Primer*. Cupertino, CA: ArcSight. Inc.

ArcSight (2005, November 29). *ArcSight ESM: Using the ArcSight Console*. Version 3.5 SP1. Cupertino, CA: ArcSight. Inc.

ArcSight (2005, December 5). *Installation and Configuration Guide*. ArcSight ESM Version 3.5. Cupertino, CA: ArcSight. Inc.

ArcSight (2006, February 9). *ArcSight 101: Concepts for ArcSight ESM*. Cupertino, CA: ArcSight. Inc.

ArcSight (2006, September 22). *ArcSight SmartConnector User's Guide: Topics Applicable to all SmartConnectors*. Cupertino, CA: ArcSight. Inc.

ArcSight (2007, March 17). *ESM System Content: Reference Guide*. ArcSight ESM v4.0. Cupertino, CA: ArcSight. Inc.

Babcock, Pete (2006). *Real World ArcSight Tuning and Customization for Actually Responding to Events and Not Just Logging Them.* ArcSight 2006 Users' Conference, Chantilly, Virginia.

Bace, Rebecca and Mell, Peter (2001). "Intrusion Detection Systems." NIST Special Publication 800-31, 19 August. Gaithersburg, MD: National Institute of Standards and Technology. Retrieved on 16 August 2007 from http://csrc.nist.gov/publications/nistpubs/800-31/sp800-31.pdf.

Barman, Scott (2002). *Writing Information Security Policies*. Indianapolis, IN: New Riders.

Bejtlich, Richard (2005). *The Tao of Network Security Monitoring: Beyond Intrusion Detection*. Upper Saddle River, NJ:

Addison-Wesley.

Bejtlich, Richard (2006). *Extrusion Detection: Security Monitoring for Internal Intrusions*. Upper Saddle River, NJ: Addison-Wesley.

DeStefano, Rocky (2007). Personal communication to Frank Olmstead, 10 July.

Denning, Dorothy E. (1987, February). "An Intrusion-Detection Model," *IEEE Transactions on Software Engineering*. Vol. SE-13, No. 2, February 1987, 222-232.

Dubrawsky, Ido and Saville, Roland (2004). *SAFE: IDS Deployment, Tuning, and Logging in Depth*. San Jose, CA: Cisco Systems, Inc.

http://www.cisco.com/warp/public/cc/so/neso/vpn/vpne/safwp_wp.pdf

Intruvert Networks, *Intrushield IDS System: Manager Administrator's Guide*. Version 1.5.b (Santa Clara, CA: Network Associates, Inc, 2003).

Kochmar, John; Allen, Julia; Alberts, Christopher; Cohen, Cory; Ford, Gary; Fraser, Barbara; et al. (1998). *Preparing to Detect Signs of Intrusion*. Security Improvement Module CME/SEI-SIM-005 Pittsburgh, PA: Software Engineering Institute, Carnegie Mellon University. Retrieved on 16 August 2007 from http://www.sei.cmu.edu/pub/documents/sims/pdf/sim005.pdf.

Koziol, Jack (2003). *Intrusion Detection with Snort*. Indianapolis, IN: SAMS Publishing. Chapter 10, Tuning and Reducing False Positives, pp. 207-232.

Martinez, Gabriel. *Innovative Approaches to Solution*

Voorhees   42

building: Essential Steps in Building Your Own Solutions by Example. ArcSight 2006 Users' Conference, Chantilly, Virginia.

Martinez, Gabriel and Veach, Al. *Use Case Education: What the Heck is a Use Case, Anyway?* ArcSight 2007 Users' Conference, Chantilly, Virginia.

Marty, Raffael (2005). *Advanced Categorization.* ArcSight 2005 Users' Conference, North Bethesda, Maryland.

Microsoft (2005). *Auditing Security Events Best Practices.* Updated January 21. 2005. Retrieved on 15 August 2005 from http://technet2.microsoft.com/WindowsServer/en/library/5658fae8-985f-48cc-b1bf-bd47dc2109161033.mspx.

Microsoft (2006). "The Security Monitoring and Attack Detection Planning Guide." Redmond, WA: Microsoft Corporation. Retrieved on 15 August 2005 from http://www.microsoft.com/technet/security/guidance/auditingandmonitoring/securitymonitoring/default.mspx.

Ranum, Marcus (2004). "IDS Tuning: At the front or back end? The case for a SIM." Faculty Columns. Institute for Applied Network Security, April. Retrieved on August 11, 2007, from http://www.ianetsec.com/news/all_fc_ranum1.htm.

RSA (2007). *Best Practices in Log Management for Security and Compliance.* White paper. Retrieved 16 August 2007 from www.rsa.com.

SANS (2007). *What Works in Event and Log Management: Driving Compliance with Log Management at Tyson Foods.* Transcript of May 31 Webcast. Email received 31 May 2007.

Saurabh, Kumar (2006). *Event Prioritization Demystified.* ArcSight 2006 Users' Conference, Chantilly, Virginia.

Stoneburner, Gary; Goguen, Alice; and Feringa, Alexis (2002). *Risk Management Guide for Information Technology Systems*. NIST Special Publication 800-30. Gaithersburg, MD: National Institute of Standards and Technology. Retrieved on 16 August 2007 from http://csrc.nist.gov/publications/nistpubs/800-30/sp800-30.pdf.

Theravukattil, Philip (2006). *Asset Modeling, Vulnerability Mapping and Use Case Development*. ArcSight 2006 Users' Conference, Chantilly, Virginia.

Thomas, Ryan (2007). *Best Practices for Content Development*. ArcSight 2007 Users' Conference, Chantilly, Virginia.

US-CERT (2006). "Federal Incident Reporting Guidelines." Retrieved on 6 September 2007 from http://www.us-cert.gov/federal/reportingRequirements.html.

Whitsitt, Jack (2007). Personal Communication, 16 September.

Wilder, Jason (2007). Personal Communication, 24 September.