



Detecting Data Loss with the Cisco Cyber Threat Defense Solution 1.0

April 9, 2012

Introduction

Of all of the potential threats facing an enterprise today, data loss is perhaps the most painful. Every company worries about having to disclose the loss of customer data or intellectual property, and facing the inevitable bad press, awkward questions, fines, and even investigations, that result. As a result, preventing persistent, widespread attacks against customer data, trade secrets, intellectual property, email, or financial data has become a top priority for every IT organization. Unfortunately detecting data loss (exfiltration) is one of the most difficult problems facing enterprise IT today.

One of the most difficult challenges for security practitioners combating data loss is determining how to gain visibility to the internal network. The internal network is where the critical data resides, and represents the first possible location where one can detect and prevent data loss. The Cisco Cyber Threat Defense Solution addresses this problem by providing the visibility mechanisms needed for detecting data loss. Cisco's solution addresses the data loss problem by combining the use of hardware-enabled NetFlow with the Cisco Identity Services Engine and a management console for inspecting flow data.

Prerequisites

This document assumes the reader has read the Cisco Cyber Threat Defense Solution Overview, Design and Implementation Guide, and the Introduction to Cisco Cyber Threat Defense "how-to" document. Readers will gain the maximum benefit from the examples in this guide if they have installed a fully functioning Cyber Threat Defense Solution, including switch and router infrastructure that is properly configured for sending NetFlow, a fully functioning Cisco Identity Services Engine environment, and a StealthWatch FlowCollector and StealthWatch Management Console. With these in place, security practitioners should then plan on following the step-by-step examples while in front of the StealthWatch console.

Solution Components

The Cisco Cyber Threat Defense Solution is composed of three integrated components:

NetFlow data generation devices. NetFlow is the de facto standard for acquiring IP operational data. Traditional IP NetFlow defines a flow as a unidirectional sequence of packets that arrive at a router on the same interface or sub-interface and have the same source IP address, destination IP address, Layer 3 or 4 protocol, TCP or UDP source port number, TCP or UDP destination port number, and type of service (ToS) byte in their TCP, UDP, and IP headers, respectively.

Flexible NetFlow is the next generation in flow technology and is a particularly valuable component of the Cisco Cyber Threat Defense Solution. Flexible NetFlow optimizes the network infrastructure, reducing operation costs and improving capacity planning and security incident detection with increased flexibility and scalability.

NetFlow can be enabled on most Cisco switches and routers, as well as some Cisco VPN and firewall devices. In addition, select devices now employ special hardware acceleration, ensuring that the NetFlow data collection process does not impact device performance. This enables NetFlow data collection pervasively throughout the network—even down to the user edge—so that every packet from every network segment and every device is completely visible.

Cisco Identity Services Engine. The Identity Services Engine delivers all the necessary identity services required by enterprise networks—AAA, profiling, posture, and guest management—in a single platform. In the context of the Cisco Cyber Threat Defense Solution, the Identity Services Engine can be deployed as either a network appliance or virtual machine and answers the “who” (user), “what” (device), and “where” (which NetFlow-enabled device) questions that tie network flow data to the actual physical network infrastructure.

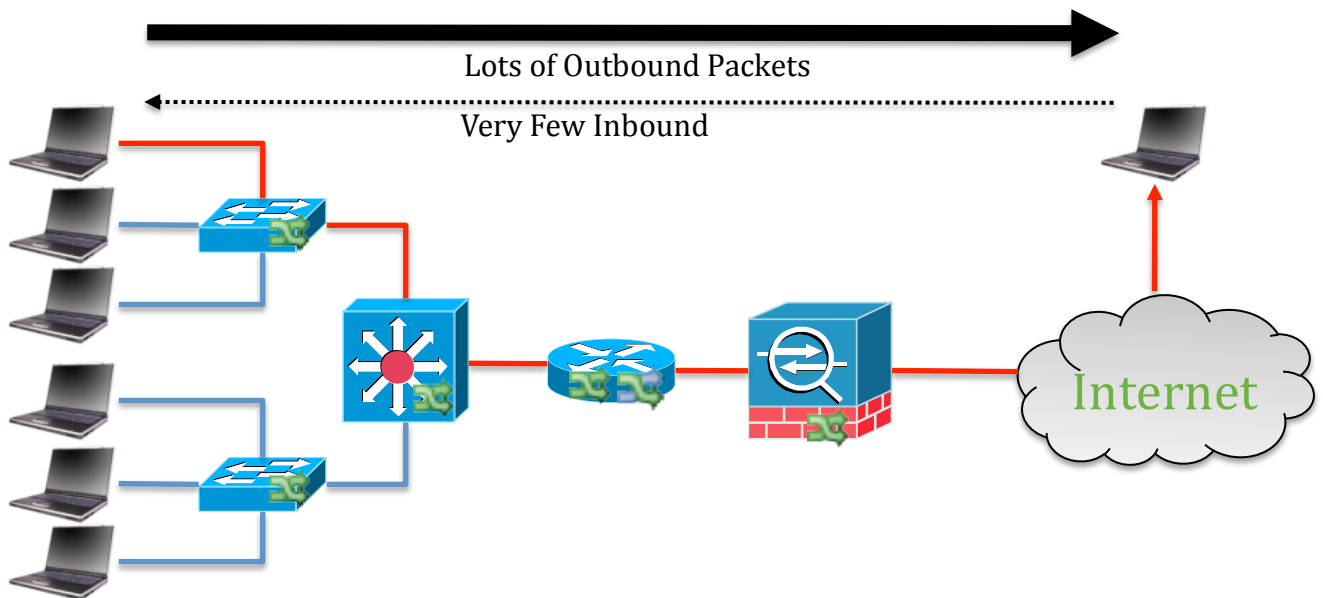
In an enterprise deployment, the Identity Services Engine provides the central policy enforcement needed to govern a network. The Identity Services Engine can provision and deliver cross-domain application and network services securely and reliably in enterprise wired, wireless, and VPN environments. This policy-based service enablement platform helps ensure corporate and regulatory compliance, enhances infrastructure security, and simplifies enterprise service operations. The Identity Services Engine can gather real-time contextual information from the network, users, and devices and make proactive governance decisions by enforcing policy across the network infrastructure.

Lancop StealthWatch System. This NetFlow visibility, network performance, and threat detection solution provides an easy-to-use interface that enables both monitoring and detailed forensics. The solution is composed of two core components: the StealthWatch Management Console and one or more StealthWatch FlowCollectors. Additional optional components include a StealthWatch FlowSensor and a StealthWatch FlowReplicator.

Operating Concepts

The Cisco Cyber Threat Defense Solution detects data loss by using a customizable set of rules within StealthWatch that detects *asymmetric outbound flows*.

An Asymmetric Flow



The rule within StealthWatch that monitors for this behavior is called *Suspect Data Loss*. The rest of this document examines how this rule operates, how to configure and tune it, and how to use it to detect data exfiltration.

Note: An asymmetric flow should not be confused with packets that follow asymmetric paths. Packets that follow asymmetric paths use *different physical or logical paths* for inbound and outbound packets. An asymmetric flow is one in which there is a great disparity between the *quantity* of inbound and outbound packets.

Determining Which Flow Traffic Is “Visible”

Within an enterprise, when two endpoints communicate across a network connected by NetFlow-enabled network access devices, the NetFlow-enabled devices generate NetFlow data that describes the characteristics of that flow. An endpoint in this network can either be inside or outside the boundaries of the corporate network.

Note: In this document, the terms “endpoint,” “computer,” “host,” and “device” are used interchangeably to mean any network-attached system.

The visibility of a flow is subsequently defined by the endpoint’s location, expressed as one of four possible “classes” of flows:

		From	
		Inside	Outside
To	Inside	Visible	Visible
	Outside	Visible	Not visible

To summarize, the Cisco Cyber Threat Defense Solution has visibility to flows that are “inside to inside,” “inside to outside,” and “outside to inside.” Because the solution is deployed inside the enterprise network, it typically would have no visibility to traffic that is “outside to outside.”

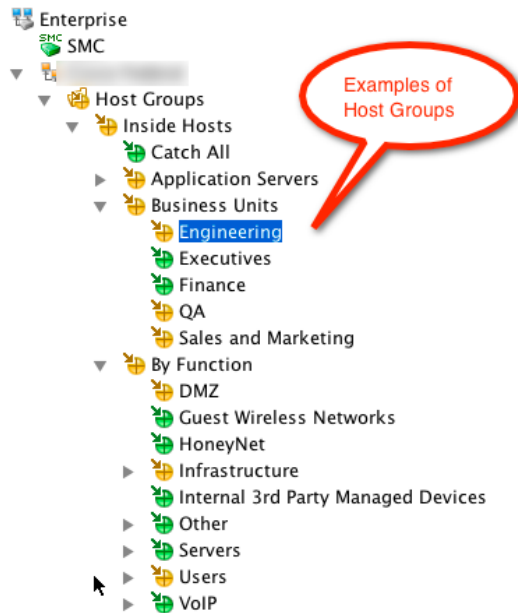
Flows Have “Direction”

Furthermore, flows have a *direction* that the Cisco Cyber Threat Defense Solution is able to discern. Consider the following example: A customer who has deployed the Cisco Cyber Threat Defense Solution in their enterprise network, “acme.com,” initiates a web browser request to <http://www.cisco.com/go/security>, an external domain. This communication will be captured by the Cisco solution as an inside-to-outside flow because it was initiated by a client inside the boundaries of acme.com and destined to a device outside of the corporate network at cisco.com.

Subsequently, clients making connections to other devices within acme.com would be classified as “inside to inside,” and any device outside the enterprise that could initiate a connection to a host inside the enterprise (for example, in a DMZ) would represent an outside-to-inside connection.

Flows Have Behavior Profiles That Can Be “Baselined”

The Cisco Cyber Threat Defense Solution captures and analyzes historical data about flows on the network and uses this to create a *baseline* of behavior. To create this baseline, a user first describes the various zones of their network using the SMC. These zones, called *host groups* in StealthWatch, can be described in many different ways to conform to an enterprise’s unique requirements. Host groups are often defined by IP address ranges and might have descriptions such as “Server Farm,” “Data Center,” “VPN IP Address Pool,” “Engineering Clients,” or “IP Phones.”



The Cisco Cyber Threat Defense Solution automatically creates *behavior profiles* of all hosts, tracking many different parameters of the flows over a period of seven days. This creates the initial behavior baseline for the host. Cisco's Cyber Threat Defense Solution automatically creates *behavior profiles* of all hosts, tracking many different parameters of the flows. The initial behavior baseline for the host is built within the first seven days of monitoring. Subsequent to those first seven days, the solution continues its baselining functions an additional 21 days to create a 28-day *rolling baseline* for the host. This baseline is used to trigger alarms when host begins to change in behavior. All 28 days of statistical information is used to trigger alarms, but the most recent 7 days are more heavily weighted.

How Data Loss Is Detected

Now, armed with visibility to three different kinds of flows—inside-inside, inside-outside, outside-inside—plus an understanding of which end initiated each flow and a baseline of each host's behavior, it is now possible to detect data loss. The solution identifies data loss by discretely measuring the quantity of data flowing in each direction and comparing it to average historical values. *Asymmetric* flows above a certain accumulated against rules until a threshold is crossed, causing the rule to trigger an alarm. The Cisco Cyber Threat Defense Solution constantly monitors these flows and compares them to baseline parameters to determine if any data is exfiltrated—all without any human intervention.

Configuration Example

The following example takes a security administrator through the process of configuring a *Suspect Data Loss* rule within the StealthWatch Management Console.

A Note About Suspect Data Loss

The Cisco Cyber Threat Defense Solution provides the ability to detect data loss based on the analysis of NetFlow data. The StealthWatch console that is used to view the NetFlow data uses a technology called a *concern index* to reflect the severity of a security event. A *concern index* is a numeric value—a counter of sorts—indicating how many times a specific kind of event has occurred outside of acceptable parameters within a window of time. The StealthWatch detection engine examines each flow as it enters the FlowCollector and then applies a set of rules to each flow. The result of the comparison between the rule and the flow data determines whether various counters should be increased.

Independent of the collection process, StealthWatch also constantly analyzes these flows to determine if thresholds have been exceeded or suspicious patterns have been detected. When a concern index value exceeds a defined threshold, StealthWatch raises an alarm, indicating a potential problem. In this document, we explore using concern indexes to detect data loss.

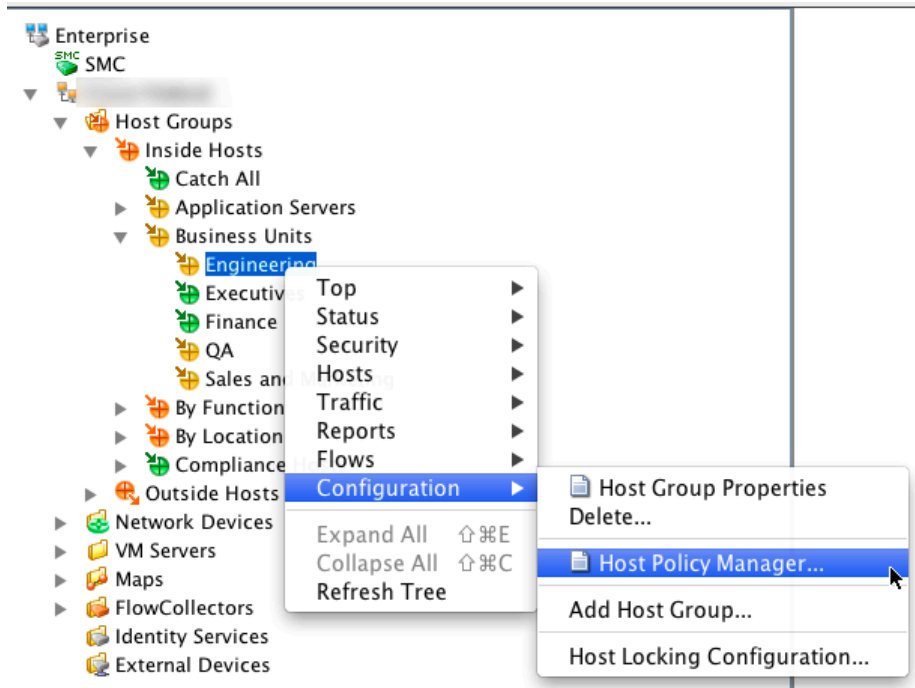
Procedure 1: Defining a Suspect Data Loss Rule

When host groups are created, they are populated with default rules. In this example, we use StealthWatch to define a new *Suspect Data Loss* rule and adjust its parameters so the reader will better understand how the Suspect Data Loss rule works.

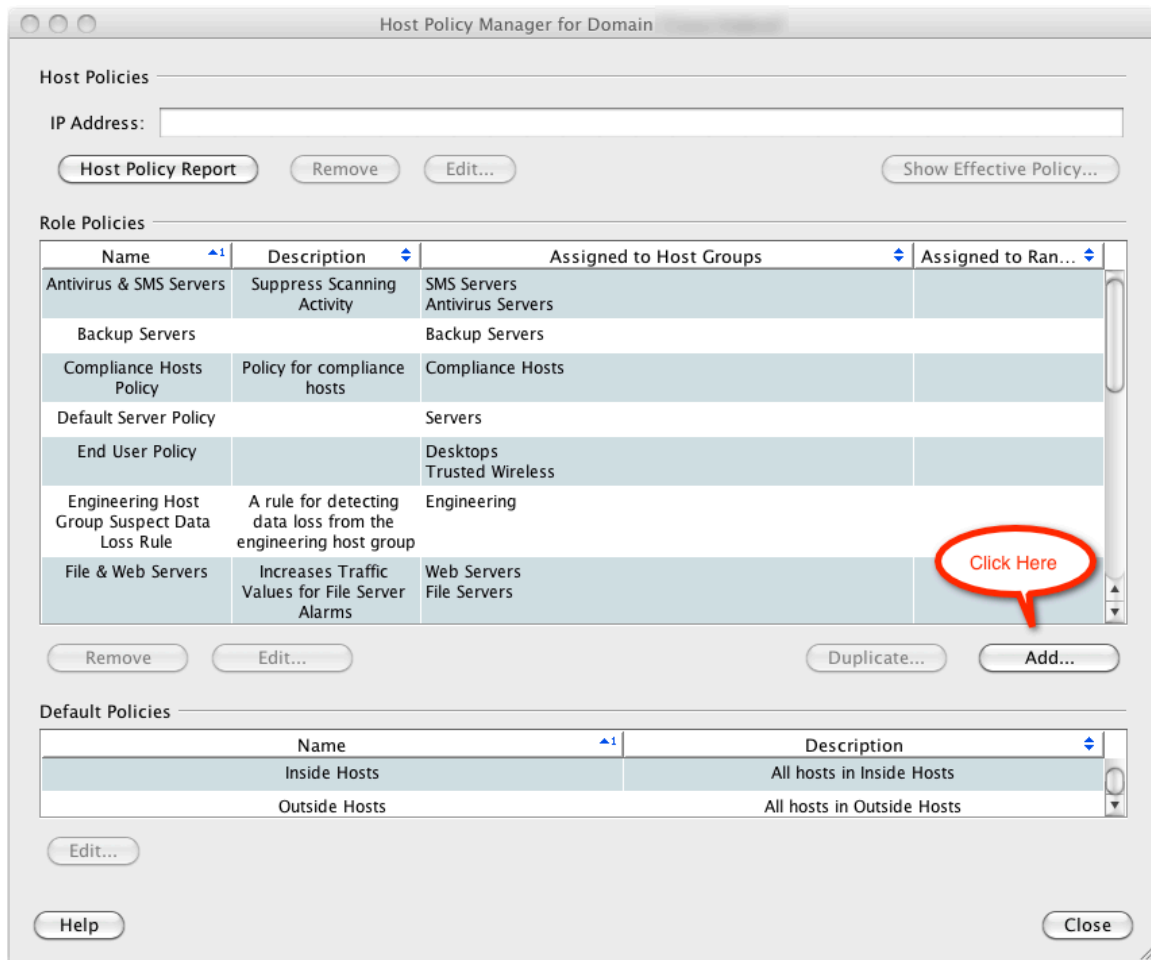
Step 1 From the main StealthWatch Management Console (SMC), navigate to the host group *Engineering*.

Step 2 Right-click the group to view the group menu.

Step 3 Select Configuration ➔ Host Policy Manager.

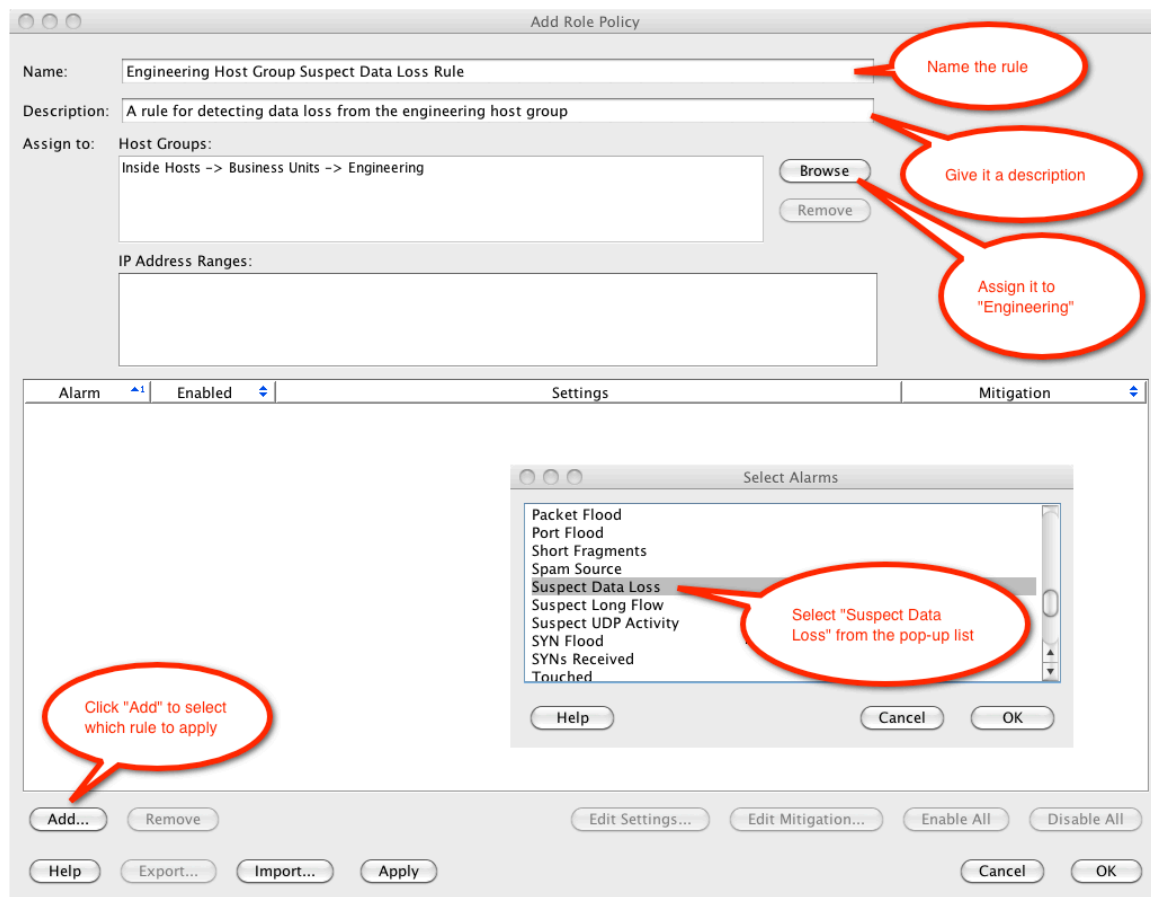


This will produce the *Host Policy Manager for Domain (Your Domain)* screen.



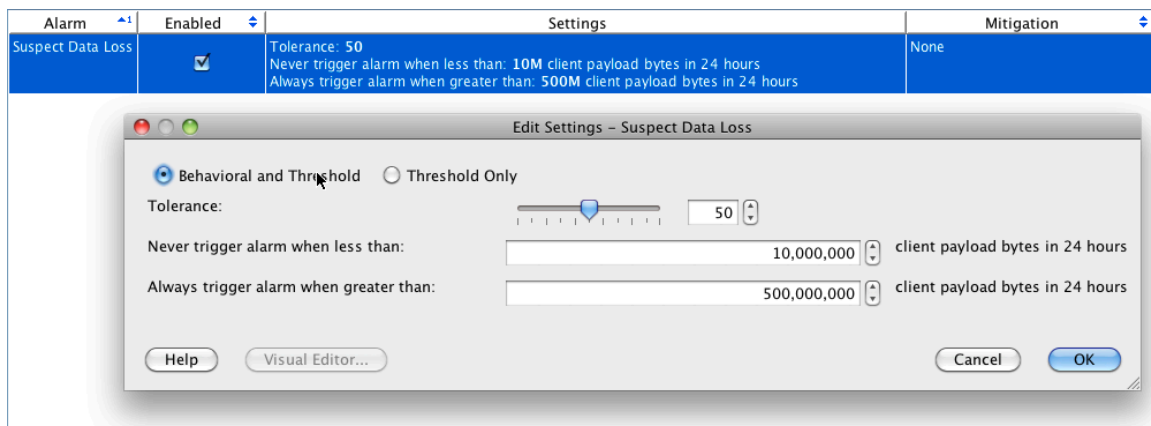
Step 4 Click the *Add* button to add a new Role Policy. This raises the *Add Role Policy* dialog box.

Step 5 Complete the fields using appropriate values for your deployment, selecting *Suspect Data Loss* for the rule.



Step 6 Click *OK* when finished. When you are done, a new *Suspect Data Loss* rule will appear.

Step 7 Right-click on this rule and select *Edit Settings*. As shown below, this dialog allows you to tune the threshold setting for this rule.

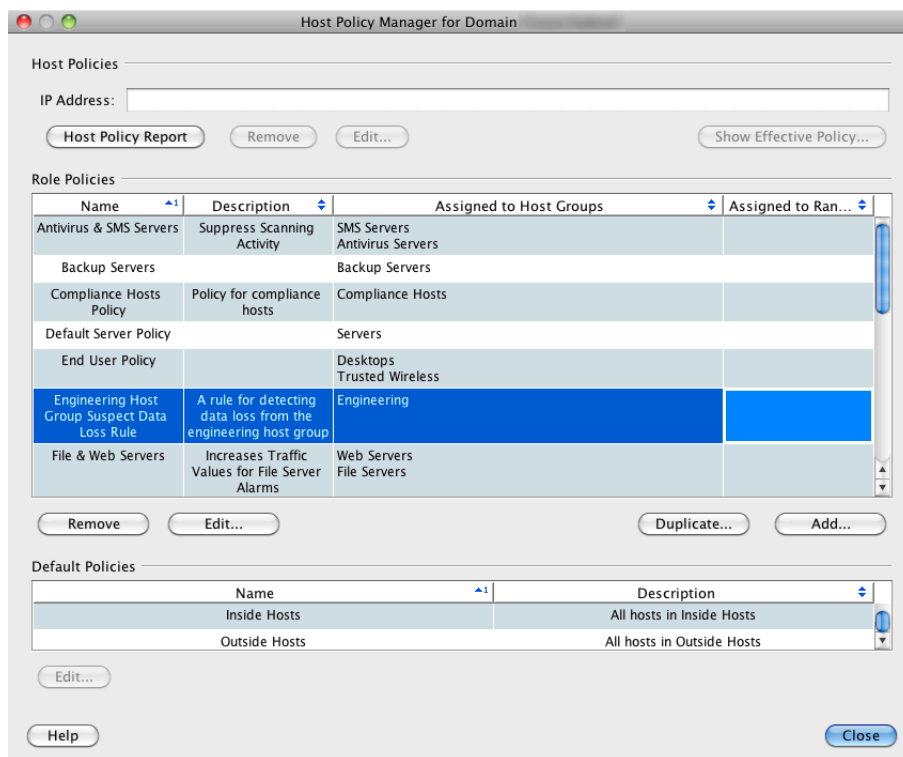


The first thing you will notice is there are two modes, *Behavioral* and *Threshold* and *Threshold Only*. *Threshold Only*-based rules trigger when the number of packets

exceeds the number defined in this dialog. *Behavior and Threshold* modifies this to allow for a low water value and a statistical variance to the threshold value defined.

The statistical variance is defined by the *Tolerance* dialog, set at “50” by default. “50” mean “fifty percent.” *Tolerance* is a relative number between 0 and 100 that indicates how much (percentage) to allow actual behavior to exceed expected behavior before raising an alarm. This allows the user to define what is “significantly different.” A tolerance of 0 means to raise an alarm for any values over the expected value; this tolerance number is very sensitive and will result in a lot of alarms. A tolerance of 100 greatly reduces the number of times an alarm is raised. With a tolerance of 50, the lowest 50% of the values over the expected value are ignored, but alarms are raised on the ones above that value. You should adjust these values to tune this rule to your specific environment. For the sake of this exercise, we leave the rule at the default of “50.”

Step 8 Click *OK* for this dialog box, and then *OK* for the *Add Role Policy* dialog as well. When you return to the *Host Policy Manager for Domain (Your Domain)* screen, you should now see your new Suspect Data Loss rule in place:



Step 9 Click *Close* to complete the process of adding this rule.

Detecting Data Loss

Now that you have an understanding of the concepts that define how data loss detection operates in StealthWatch and have created a new Suspect Data Loss rule, let's explore how data loss rules work in a production environment.

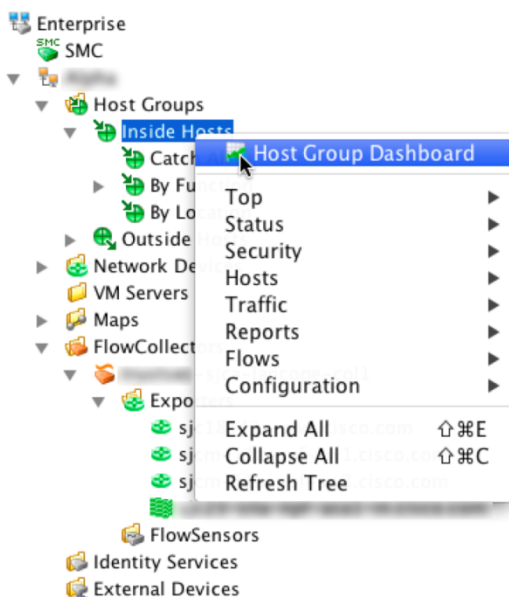
Find a Suspect Data Loss Alarm

The first step in using this new rule is to find a live Suspect Data Loss event within the network.

Step 1 Select a high-level host group, such as *Inside Hosts*.

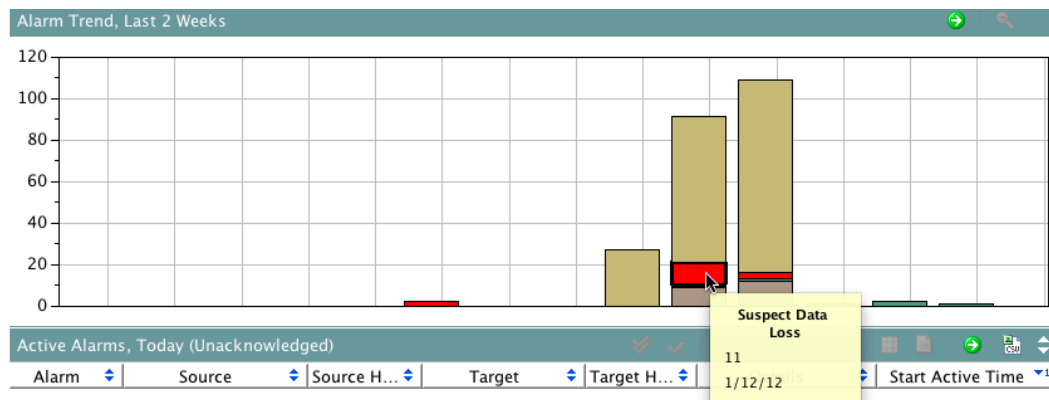
Step 2 Right-click on *Inside Hosts* to create the popup.

Step 3 Select *Host Group Dashboard* from the menu.



The result is a screen that contains a graph in the upper-right corner called the *Alarm Trend* graph. If you hover your mouse over any of the components in the graph, it will tell you what caused the event.

Step 4 Hover over the elements until you find a Suspect Data Loss event in the graph, as shown below:



At this point, you have located a Suspect Data Loss event using the graph. In the next step, we drill down into this data to learn more about the data loss event.

Filter Your Data

One of the most important processes you need to learn in StealthWatch is how to filter data. Often, the result of a query will be either too much or too little data based on what you are exploring. To adjust the result, you will need to apply filters to your data.

Step 1 Double-click on a Suspect Data Loss event. This will produce a table summarizing the events described in the graph. This could result in a large table of alarms, no alarms at all, or something in between, depending on the filter conditions applied to the data. The key to showing the records that generated the graph data is to properly select (or de-select) the filter conditions for this table, as described below.

Step 2 Click on the *Filter* button in the upper-left corner of this table.

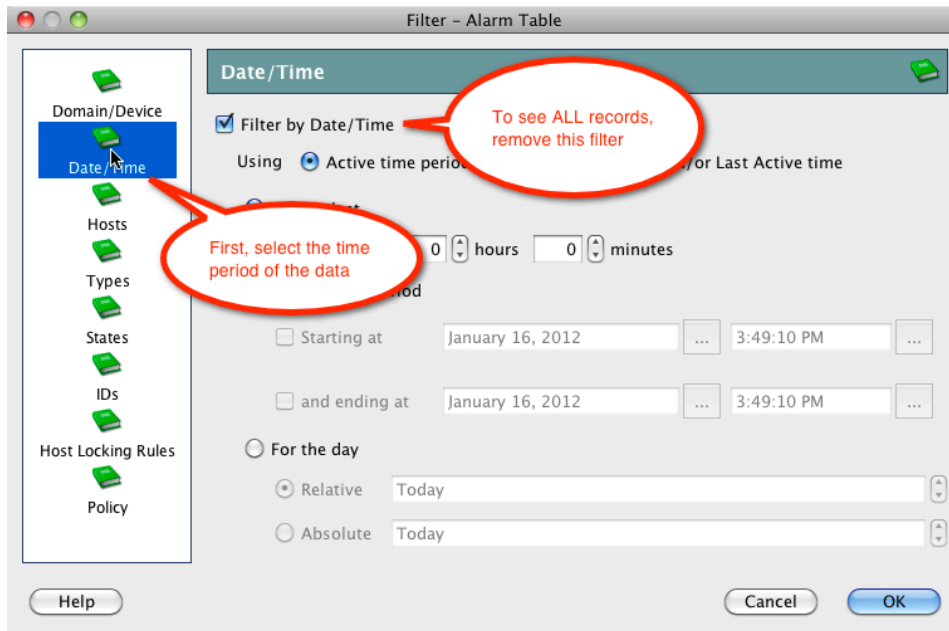
The screenshot shows the "Alarm Trend" table with the "Filter" button circled in red. The table has columns: Alarm, Source, Source Host Group, Source User, and Target. The first row is highlighted in blue and shows a "Suspect Data Loss" event on Jan 7, 2012, at 8:50:00 PM, with source 10.33.25.254 and target "Catch All". The second row is highlighted in light blue and shows a "Suspect Data Loss" event on Jan 8, 2012, at 3:15:00 PM, with source 10.33.25.254 and target "Catch All". The third row is highlighted in light blue and shows a "New Host Active" event on Jan 11, 2012, at 12:10:00 PM, with source 10.28.133.104 and target "10.34".

Alarm	Source	Source Host Group	Source User	Target
Suspect Data Loss	10.33.25.254	Catch All		Multiple
Suspect Data Loss	10.33.25.254	Catch All		Multiple
New Host Active	10.28.133.104	Catch All		10.34

When you click on the *Filter* button, it produces a dialog box. On the left side of this dialog box is a list of filter conditions expressed by green book icons. We need to click on some of these icons to ensure we have the proper filter conditions in place for this table.

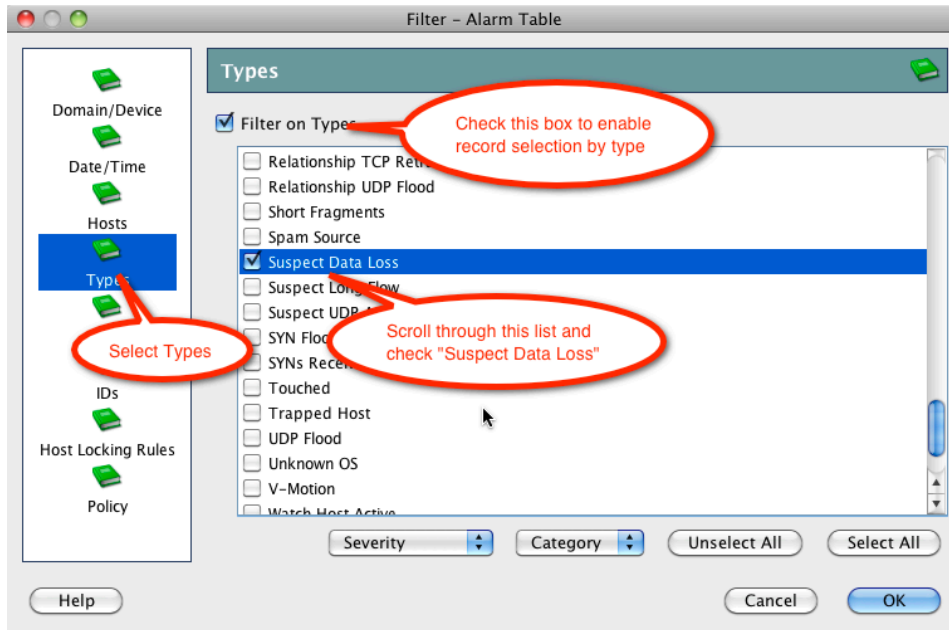
Step 3 Select *Date/Time*. Depending on how much data you have, you may wish to filter this data by a date and time range.

Step 4 Click *Filter by Date/Time* to filter the data by date and time. In this case, however, we've elected NOT to filter using date and time, so we uncheck this box. This means no date or time filter will be applied to the data.



Step 5 Next, select *Types* from the filter list. This allows us to filter the data based on the type of alarm we are interested in exploring.

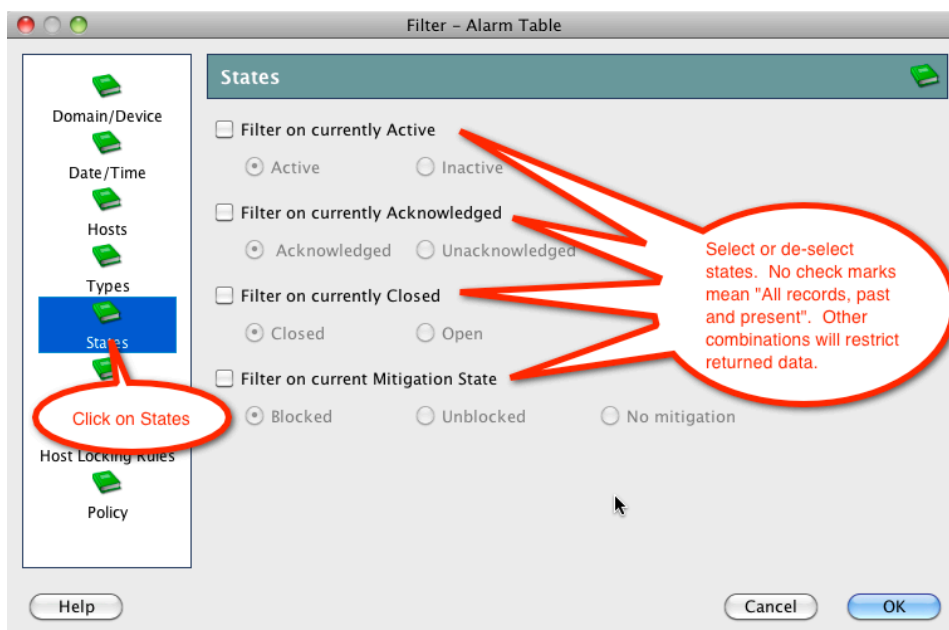
Step 6 Click on *Filter on Types*, and scroll through the list until you see *Suspect Data Loss*. Make sure this is selected.



Finally, you need to select the *state* (condition) of the alarm. This is perhaps the most important filter you need to set, and one that is easy to overlook.

Step 7 Select *States*. There are four conditions that can be set: *Active*, *Acknowledged*, *Closed*, and *Mitigation State*. Since we want to see *all* records, we uncheck all of these.

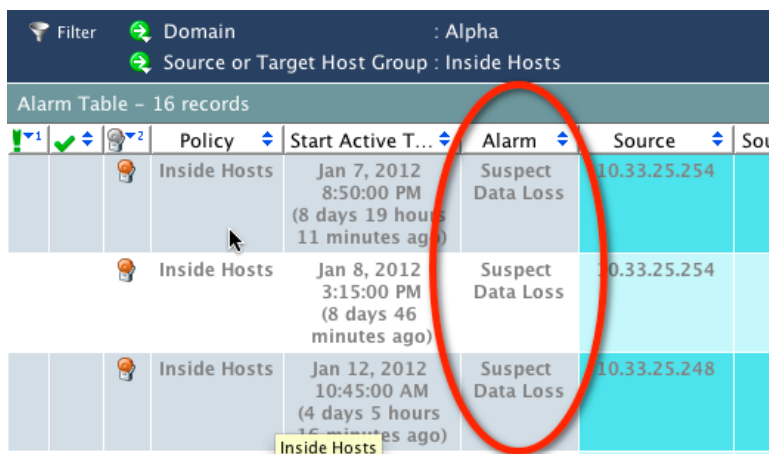
Step 8 Uncheck all of the boxes. If you were interested in alarms in only one state (for example, *Currently Active* alarms) you would check that box.



Once you've completed this last step, you have finished setting all the filter conditions.

Step 9 Click *OK*.

The screen will now return you to your original table, filtered by the conditions you've just set. As seen below, you should now have a table populated solely by Suspect Data Loss alarms.



Filter Domain : Alpha
Source or Target Host Group : Inside Hosts

Alarm Table - 16 records

			Policy	Start Active T...	Alarm	Source	Sou
			Inside Hosts	Jan 7, 2012 8:50:00 PM (8 days 19 hours 11 minutes ago)	Suspect Data Loss	10.33.25.254	
			Inside Hosts	Jan 8, 2012 3:15:00 PM (8 days 46 minutes ago)	Suspect Data Loss	10.33.25.254	
			Inside Hosts	Jan 12, 2012 10:45:00 AM (4 days 5 hours 16 minutes ago)	Suspect Data Loss	10.33.25.248	

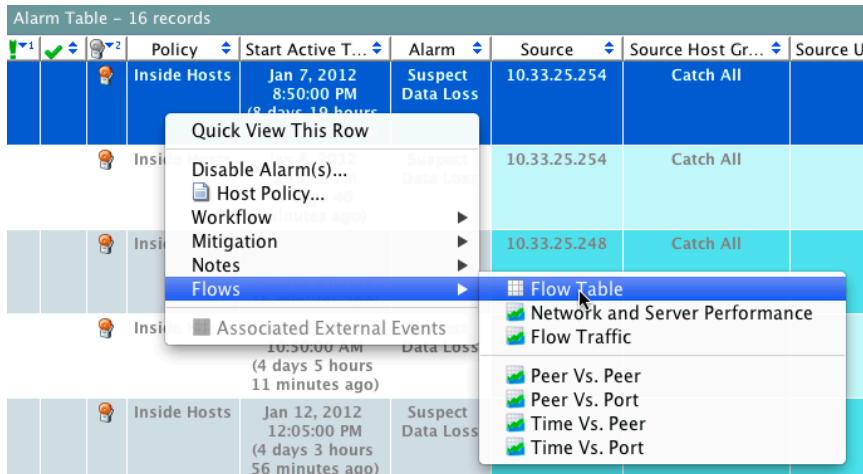
Investigating the Alarm and Identifying the Responsible User

Now that we have filtered our data to only show Suspect Data Loss events, we want to select one of these events to show us the underlying flows that generated the event. That data is expressed in a flow table, and might represent numerous flows.

Step 1 Select a row you want to explore.

Step 2 Right-click on that row.

Step 3 Select Flows → Flow Table from the resulting popup menu.



This will result in a table with one or more records representing the flows that generated the graph data. From this table, you are looking for records that show a wide disparity between the quantities of data sent outbound by the client to a destination. This data is expressed in the table in the column labeled *Client Ratio (%)*. In the following table, we've selected a good example. Notice that the client has sent 1.11 GB of data outbound to another device that has only responded with 12.56 MB inbound. The ratio is 98.91% outbound, and the quantity of data is quite substantial.

Total Bytes ▾	Total Packets ▴	Start Active Time ▴	Client Bytes ▴	Client Ratio (%) ▴	Server Bytes ▴
1.12G	964,396	Jan 7, 2012 8:26:08 PM (8 days 20 hours 1 minute ago)	1.11G	98.91% <div></div>	12.56M

Step 4 Select this row from the table, making sure to click on the IP address in the *Client Host* column.

Step 5 Right-click on the row.

Step 6 Select *Host Snapshot* from the resulting popup menu.

Flow Table - 139 records

Client Host	Client Host Groups
10.33.25.254	Catch All
10.33.25.254	Quick View This Row
10.33.25.254	Add to Short List
10.33.25.254	Remove from Short List
10.33.25.254	Replace Short List
10.33.25.254	Host Peer Chart
10.33.25.254	Host Port Chart
10.33.25.254	for Host 10.33.25.254:
10.33.25.254	Host Snapshot
10.33.25.254	Top
10.33.25.254	Status
10.33.25.254	Security
10.33.25.254	Hosts
10.33.25.254	Traffic
10.33.25.254	Reports
10.33.25.254	Flows
10.33.25.254	Configuration
10.33.25.254	External Lookup

This will produce the *Identity and Device Table*, showing the Cisco Identity Services Engine device the user authenticated against, the user name of the user, and other associated information that now makes it a straightforward task to locate the device responsible for the data loss.

Identification	Alarms	Security	CI Events	Top Active Flows	Identity, DHCP & Host Notes	Exporter Interfaces
Identity and Device Table - 1 record						
Start Active Time	End Active Time	Cisco ISE	User Name	MAC Address	Identity Group	
Dec 20, 2011 4:26:38 PM (15 days 18 hours 50 minutes ago)	Current	DemoISE (172.29.5.39)	user1	00:50:56:90:00:98 (VMware, Inc.)	demousers,Profiled	

Conclusion

Data loss is a significant problem facing most companies today. Detecting data loss is difficult because networks do not traditionally provide good visibility to data exfiltration. The Cisco Cyber Threat Defense Solution addresses this problem by providing the visibility mechanisms needed for detecting data loss. Cisco's solution integrates the Lancope StealthWatch System with Cisco's hardware-accelerated NetFlow and the Identity Services Engine to provide a convenient and effective way to address the problem of data loss.