IBM Proventia Content Analyzer Guidelines

December 11, 2007

Overview

Introduction

Proventia Content Analyzer uses the data inspection and analysis capabilities in Proventia Network Intrusion Prevention (IPS) appliances to find Personal Identifiable Information (Pii) or other confidential information moving through and out of your network.

Content Analyzer can detect both accidental and malicious data leakage. It can be a powerful tool in your data leakage prevention (DLP) strategy.

How it works

Content Analyzer inspects data packets as they move across the network, detecting the transmission of many types of confidential information. Content Analyzer can identify patterns such as credit card numbers, names, dates, dollar amounts, e-mail addresses, social security numbers, US phone numbers, and US postal addresses in various protocols and content.

In addition to the preset signatures, you can create up to eight custom *user-defined* signatures. You can also create up to eight *user-combined* signatures by grouping combinations of preset and user-defined signatures. A user-combined signature functions as a single dataset.

Supported agents

Content Analyzer currently works with Proventia Network Intrusion Prevention (IPS) appliances. It is supported for any firmware version that has not yet reached end-of-content (EOC). Other agents may be supported, even though they are not specifically described in these guidelines.

Required configuration

Content Analyzer is delivered in an XPU for Proventia Network IPS appliances, but it is disabled by default.

You must do the following before you can use Content Analyzer:

● Enable the signatures you want to use (on the Security Events page)

● Enable Content Analyzer and set the parameters you want to use (on the Global Tuning Parameters or the Local Tuning Parameters page)

Important Considerations

Introduction

This section contains points to consider as you evaluate whether Proventia Content Analyzer is right for your environment.

Performance and tuning

With all Content Analyzer signatures and protocols turned on, you may notice some impact to network performance. Few enterprises need this level of protection, and your performance numbers are likely to improve as you identify the subset of signatures and protocols you need.

You may see a large number of events based on certain signatures and content types. You can reduce the number of events by thoroughly tuning your Content Analyzer policy.

Note: If you need assistance with tuning your policies, our professional security consultants are available to help.

Setting boundaries

Consider deploying and enabling Content Analyzer in audit mode on segments where confidential information should be encrypted at all times, such as perimeter or DMZ boundaries.

Monitoring or blocking

You can use Content Analyzer for either auditing or blocking.

Most enterprises use audit mode while they are tuning policies. This approach helps security managers understand the kinds of data that they *could* be blocking without disrupting business operations. Other enterprises find that audit mode is sufficient, and they have no plans to deploy in blocking mode.

Supported protocols and file formats

Overview

Proventia Content Analyzer can inspect a variety of protocols and file formats, as outlined in this topic.

Supported protocols

Proventia Content Analyzer supports the following communication protocols:

- HTTP
- FTP
- SMB
- SMTP
- IMAP
- POP3
- Microsoft Messenger
- Yahoo Messenger
- AOL Messenger
- IRC

For messaging and e-mail protocols (such as SMTP, IMAP, POP3, and chat), Content Analyzer evaluates both inline text and attachments.

Supported file formats

Proventia Content Analyzer supports the following file formats:

- Compound document file format (for Microsoft Office documents)
- Portable Document Format (PDF)
- Text (TXT, CSV)
- Rich Text (RTF)
- Extensible Markup Language (XML)
- Hypertext Markup Language (HTML)
- ZIP (Deflate format)
- gzip

Signatures for Proventia Content Analyzer

Introduction

Proventia Content Analyzer provides three categories of signatures:

- Preset signatures
- User-defined signatures
- User-combined signatures

Preset signatures

Proventia Content Analyzer offers eight preset signatures that you can use to identify common types of data leakage.

Signature	Description
Content_Analyzer_Credit_Card_Num	Looks for patterns that are typical for credit card numbers
Content_Analyzer_Postal_Addr	Looks for patterns that are typical for physical or mailing addresses
Content_Analyzer_Social_Security_Num	Looks for patterns that are typical for social security numbers
Content_Analyzer_Dollar_Amount	Looks for patterns that are typical for financial data
Content_Analyzer_Date	Looks for patterns that are typical for various date formats
Content_Analyzer_US_Phone_Num	Looks for patterns that are typical for phone numbers
Content_Analyzer_Email_Addr	Looks for patterns that are typical for e-mail addresses
Content_Analyzer_Person_Name	Looks for patterns that are typical for various name formats

Table 1: *Preset signatures for Proventia Content Analyzer*

User-defined signatures

In addition to preset signatures, Content Analyzer lets you define up to eight custom signatures to meet your needs. For example, you can create a signature to recognize the format you use for account numbers or policy numbers.

You can define the following user-defined signatures through Global Tuning Parameters:

- Content_Analyzer_User_Defined_0
- Content_Analyzer_User_Defined_1
- Content_Analyzer_User_Defined_2
- Content_Analyzer_User_Defined_3
- Content_Analyzer_User_Defined_4
- Content_Analyzer_User_Defined_5
- Content_Analyzer_User_Defined_6
- Content_Analyzer_User_Defined_7

User-combined signatures

You can combine signatures to create custom combinations. For example, you might not be concerned about a single unencrypted postal address going out in an e-mail message. However, an e-mail containing a postal address combined with a name and social security number might be cause for concern.

You can combine preset signatures, user-defined signatures, or both preset and user-defined signatures to create conjoined datasets to be used to analyze traffic.

You can define the following user-combined signatures through Global Tuning Parameters:

● Content_Analyzer_User_Combined_0

● Content_Analyzer_User_Combined_1

● Content_Analyzer_User_Combined_2

● Content_Analyzer_User_Combined_3

● Content_Analyzer_User_Combined_4

● Content_Analyzer_User_Combined_5

● Content_Analyzer_User_Combined_6

● Content_Analyzer_User_Combined_7

Enabling signatures

To enable Content Analyzer signatures:

1. From the Security Events page, sort by **XPU 27.120** to isolate the Content Analyzer signatures.

2. Select the Content Analyzer signatures you want to enable.

Enabling Content Analyzer and Setting Parameters

Introduction   Proventia Content Analyzer is disabled by default. To use this feature, you must enable it. You can enable Content Analyzer for an appliance using Proventia Manager or for multiple appliances using SiteProtector.

Use the **Global Tuning Parameters** page to enable Content Analyzer.

Enabling Content Analyzer   To enable Proventia Content Analyzer:

1. Select **Intrusion Prevention→Global Tuning Parameters**.

2. Click the Add   icon.

3. Use the following information to complete the fields:

Field	Your entry
Name	**pam.ca.enabled**
Value	true

Setting parameters   To set parameters for Content Analyzer:

1. Select **Intrusion Prevention→Global Tuning Parameters**.

2. Click the Add   icon.

3. Type the parameter name and value.

Tuning Parameters

Overview Content Analyzer includes two classes of tuning parameters:

- Module parameters that control how Content Analyzer operates
- Parameters that are specific to security checks

Module parameters Use the following tuning parameters to control how Content Analyzer functions in your environment:

Parameter	Description
pam.ca.enabled	Enables or disables Content Analyzer ● **True**. Enables Content Analyzer ● **False**. Disables Content Analyzer (Default)
pam.ca.aolimft.enabled pam.ca.aim.enabled pam.ca.ftp.enabled pam.ca.http.enabled pam.ca.imap4.enabled pam.ca.irc.enabled pam.ca.ircft.enabled pam.ca.msmsgr.enabled pam.ca.msmsgrft.enabled pam.ca.pop.enabled pam.ca.smb.enabled pam.ca.smtp.enabled pam.ca.yahoo.enabled pam.ca.yahooft.enabled	Controls which communication protocols are inspected by Content Analyzer ● **True**. Enables checking for the specified protocol (Default) ● **False**. Disables checking for the specified protocol **Note:** By default, all protocols are enabled for checking.
pam.ca.report.packetinfo	Controls whether Content Analyzer displays the packet information (containing sensitive information) as part of the event details ● **True**. Displays all information (Default) ● **False**. Masks confidential information
pam.ca.uncompress.enabled	Controls whether Content Analyzer uncompresses ZIP files and inspects their contents ● **True**. Uncompresses and inspects ZIP files (Default) ● **False**. Does not inspect ZIP files **Note:** Uncompressing ZIP files offers additional protection, but can have a negative effect on performance.

Table 2: *General tuning parameters for Content Analyzer*

Contents of document subject to change.

Parameters for customizing signatures

Other parameters control settings that are specific to a particular security check. You can use these parameters to set the following options for signatures:

- Regular expressions
- Minimum number of required matches
- Validation for credit card numbers (for credit card signature only)
- Identifying text to be reported as part of a user-defined event
- Identifying text to be reported as part of a user-combined event

Regular expressions

You can define specific values for which Content Analyzer should trigger events. Use these parameters to set values that are specific to your organization.

Use the following parameters to set regular expressions for signatures:

Parameter	Description
pam.ca.credit_card_num.regex pam.ca.postal_addr.regex pam.ca.social_security_num.regex pam.ca.dollar_amount.regex pam.ca.date.regex pam.ca.us_phone_num.regex pam.ca.email_addr.regex pam.ca.person_name.regex	Sets specific values that Content Analyzer should look for. **Example:** If you set pam.ca.date.regex to "2008," any occurrence of "2008" triggers an event. The following strings would trigger an event: • 2008 • 770-555-2008 • 2008 Main Street

Table 3: *Parameters for regular expressions*

Example: A security manager might not be concerned with most credit card numbers that are transmitted over the network. However, the security manager could set up a regular expression to trigger an event if an employee tries to send a specific credit card number (such as the corporate credit card number) through the network.

Minimum number of required matches

You can set the minimum number of matches required to cause an event. Use the following parameters to set the minimum number of matches for specific signatures:

Parameter	Description
pam.ca.credit_card_num.minmatch pam.ca.postal_addr.minmatch pam.ca.social_security_num.minmatch pam.ca.dollar_amount.minmatch pam.ca.date.minmatch pam.ca.us_phone_num.minmatch pam.ca.email_addr.minmatch pam.ca.person_name.minmatch	Defines the minimum number of matches required to trigger an event **Default values:** social_security_num.minmatch=10 pam.ca.credit_card_num.minmatch=10 Default value for all others is 100.

Table 4: *Parameters for setting minimum number of required matches*

Example: A security manager might not want to be notified if one credit card number goes out in an e-mail message. However, five or more credit card numbers in an outgoing message might be cause for concern. In this case, the security manager would set the value for the minimum number of matches to "five."

Credit card validation

You can set Content Analyzer to verify that a number uses a valid credit card format before triggering an event. Use the following parameter to avoid reporting numbers that may resemble credit card numbers, but do not pass Mod 10 validation:

Parameter	Description
pam.ca.credit_card_num.validate	Controls whether Content Analyzer uses Mod 10 validation to verify credit card numbers ● **True**. Enabled Mod 10 validation (Default) ● **False**. Disables Mod 10 validation

Table 5: *Tuning parameter for credit card validation*

User-defined and user-combined signatures

Use tuning parameters to create user-defined and user-combined signatures. See "User-Defined Signatures" on page 11 or "User-Combined Signatures" on page 12.

Identifying text for user-defined events

You can define a text string to help identify user-defined events in the event log. Use the following parameters to assign identifying text for user-defined events:

Parameter	Description
pam.ca.user_def_0.reportstr pam.ca.user_def_1.reportstr pam.ca.user_def_2.reportstr pam.ca.user_def_3.reportstr pam.ca.user_def_4.reportstr pam.ca.user_def_5.reportstr pam.ca.user_def_6.reportstr pam.ca.user_def_7.reportstr	Defines the text string to be reported with the user-defined signature Example: "InsurancePolicyNumber"

Table 6: *Parameters for defining text string for user-defined events*

Identifying text for user-combined events

You can define a text string to help identify user-combined events in the event log. Use the following parameters to assign identifying text for user-combined events:

Parameter	Description
pam.ca.user_comb_0.reportstr pam.ca.user_comb_1.reportstr pam.ca.user_comb_2.reportstr pam.ca.user_comb_3.reportstr pam.ca.user_comb_4.reportstr pam.ca.user_comb_5.reportstr pam.ca.user_comb_6.reportstr pam.ca.user_comb_7.reportstr	Defines the text string to be reported with the user-combined signature Example: "Name/Address/PhoneNumber"

Table 7: *Parameters for defining text string for user-combined events*

User-Defined Signatures

Overview

You can configure up to eight user-defined signatures. You can configure these signatures to trigger alerts based on text that you specify using POSIX extended regular expressions. For example, a security manager could create a user-defined signature to trigger events based on the format the company uses for account numbers.

You can set up user-defined events from either the Global Tuning Parameters page or the Local Tuning Parameters page.

Available user-defined signatures

You can configure the following user-defined signatures:

- Content_Analyzer_User_Defined_0
- Content_Analyzer_User_Defined_1
- Content_Analyzer_User_Defined_2
- Content_Analyzer_User_Defined_3
- Content_Analyzer_User_Defined_4
- Content_Analyzer_User_Defined_5
- Content_Analyzer_User_Defined_6
- Content_Analyzer_User_Defined_7

Configuring user-defined signatures

To configure a user-defined signature:

1. Decide which user-defined parameter (0-7) you want to configure.

2. From the tuning parameters page, click the Add icon.

3. Type the parameter name and the value you want to assign to it.

Assigning a text string to appear in the event log

You can use parameters to assign an identifying text string to user-defined events. For example, "Content_Analyzer_User_Defined_1" in the event log might not provide enough information. A security manager can assign a text string, such as "InsurancePolicyNumber," to appear as part of the event.

References

Refer to the *IBM Proventia Network Intrusion Prevention System G/GX User Guide* for detailed information about creating regular expressions.

Contents of document subject to change.

User-Combined Signatures

Overview

You can combine preset signatures, user-defined signatures, or both preset and user-defined signatures to create *user-combined* signatures. A user-combined signature is a conjoined dataset that functions as a single check. A parameter defines the list of Content Analyzer signatures to be included in a user-combined signature.

Often it is the *combination* of data that makes it confidential. For example, a dollar amount without any other information could be harmless. A dollar amount with a person's name might indicate that someone is transmitting salary information.

You can set up user-defined events from either the Global Tuning Parameters page or the Local Tuning Parameters page.

Available user-combined signatures

You can define the following user-combined signatures:

- Content_Analyzer_User_Combined_0
- Content_Analyzer_User_Combined_1
- Content_Analyzer_User_Combined_2
- Content_Analyzer_User_Combined_3
- Content_Analyzer_User_Combined_4
- Content_Analyzer_User_Combined_5
- Content_Analyzer_User_Combined_6
- Content_Analyzer_User_Combined_7

Guidelines and syntax

Keep the following guidelines in mind as you define user-combined signatures:

- You can use preset signatures, user-defined signatures, or both to create user-combined signatures.
- You can use the **issue name** or **issue id**.
- Use a space ( ) to separate signatures.
- An exclamation point (!) functions as a "not" statement. It instructs Content Analyzer to exclude the value that follows it.
- Individual signatures *do not* have to be enabled to be used as part of a user-combined signature.

Example

User-combined signature:
pam.ca.user_comb_0.events=Content_Analyzer_Social_Security_Num
Content_Analyzer_Person_Name ! Content_Analyzer_User_Defined_0

Result:
This user-combined signature would trigger an event if both
"Content_Analyzer_Social_Security_Num" and "Content_Analyzer_Person_Name"
matched, but "Content_Analyzer_User_Defined_0" did not.

Configuring user-combined signatures

To configure a user-combined signature:

1. Decide which user-combined parameter (0-7) you want to configure.

2. From the tuning parameters page, click the Add icon.

3. Type the parameter name and the value you want to assign to it.

Assigning a text string to appear in the event log

You can use parameters to assign an identifying text string to user-combined events. For example, "Content_Analyzer_User_Combined_1" in the event log might not provide enough information. A security manager can assign a text string, such as "Name/Address/PhoneNumber," to appear as part of the event.

Contents of document subject to change.