



IBM Proventia® Content Analyzer Technology

*Hybrid Protection (System & Data Security) for
IBM Proventia Network Intrusion Prevention System*

*Author: John W. Pirc (NSA-IAM, CEH)
Sr. Product Manager*

Contents

2	Introduction
3	Data Security - Content vs. Context
4	<i>Content Only Scenario</i>
4	<i>Content with Context Scenario</i>
5	Protecting Content with Proventia Content Analyzer Technology
7	<i>Deployment Considerations</i>
8	<i>Getting Started – A Primer on Proventia Content Analyzer Functionality</i>
9	<i>Determining a Baseline Configuration</i>
10	<i>Proventia Content Analyzer Signature Decision Process</i>
11	<i>Use Case Deployment Scenario</i>
15	<i>Additional Use Cases</i>
18	Conclusion

Introduction

Secure and appropriate transfer of sensitive data is paramount to complying with today's data security regulations and to protecting confidential corporate data from misuse and theft. Determining whether or not the outbound data is in violation of these regulations and corporate policy can be a significant challenge. A data content analysis solution, delivered through network intrusion prevention system technology, can provide the unique benefit of inspecting both inbound and outbound data for various content patterns in support of a comprehensive data security strategy. In addition, it also protects the network from evolving vulnerabilities and exploits. As organizations implement various and layered solutions to protect data, network content analysis becomes a critical component for monitoring success, identifying weakness and providing stopgaps.

IBM Proventia® Content Analyzer technology, existing as part of the IBM Proventia Network Intrusion Prevention System (IPS), provides additional network traffic analysis that organizations can leverage to provide targeted visibility into data security issues that can be addressed with content pattern recognition. Proventia Content Analyzer inspects and identifies sensitive data through the use of detection signatures for some of the most common types of at-risk sensitive data, such as Personally Identifiable Information (PII), including credit card and social security numbers, e-mail addresses, names, dates and other data formats that are specific to the U.S., such as postal addresses and currency amounts. Users can extend this functionality into organizational or region-specific areas by constructing additional search criteria using regular expressions either singularly, or in combination with other signatures. By identifying apparent violations of corporate and regulatory policies, Proventia Content Analyzer can help organizations to recognize, track, and block traffic associated with major categories of data security risk and reveal broken business processes.

Addressing the increasing risks associated with data security and compliance requires awareness of the levels, frequency and type of inappropriate data flows. The following white paper examines how an organization can utilize existing IBM infrastructure technology to augment a data security strategy and provide additional insight into protection against malicious and inadvertent misuse of data. These insights and protections enhance the ability to evaluate and optimize multi-layered defenses across an enterprise. This paper also examines how to tune Proventia Content Analyzer to maximize the technology's detection capability.

Data Security – Content vs. Context

Data security requires a broad set of disciplines and multiple layers of technology to ensure that the variety of risks – internal and external – are mitigated. Before an organization can develop a strategy for protecting its data, it is important to understand the information life cycle and its impact on the organization.

The electronic information life cycle starts the moment content is created, even prior to writing it to media, and ends when that specific content is properly destroyed. During the lifetime of that content, it will likely be transmitted, encrypted, stored, printed, etc., any number of times. The management and protection of that information for most large organizations can be a daunting task, and the question is often asked, “Where do I start?” Data security begins with understanding the particular organizational infrastructure, corporate policy, overarching regulatory compliance requirements and most importantly, understanding what data, if released outside corporate control, might present a risk and liability.

The Payment Card Industry Data Security Standard (PCI-DSS) is one of the most recognized security regulations today and continues to evolve with even tighter requirements for protecting PII. Thanks to the penalties associated with noncompliance, unencrypted outbound PII or confidential information comprising either single or compound submissions, poses a substantial risk to both an organization's customer base and its bottom line.

When dissecting an organization's data security needs, two considerations are important in determining the breadth and depth of the organization's requirements; content and context. Content refers to the actual data itself – the PII or other sensitive information at risk of being divulged. Context refers to the situation in which the content is being used. Context often defines the degree to which sensitive content is at risk. The following scenario explains the relationship between content and context and helps explain how Proventia Content Analyzer fits into the discussion between the two.

Content Only Scenario

A 401k worksheet containing employee names, addresses and social security numbers is detected leaving the organization's perimeter unencrypted. With Proventia Content Analyzer configured on the network IPS, the MSWord document is opened and it is determined that PII is contained within the document. The capabilities of Proventia Content Analyzer allow the organization to recognize the transmission of this sensitive data and then block, generate an alert or do both.

Content with Context Scenario

After the document was opened and the PII identified, an organization's larger data loss prevention solution can help identify the source of the original transmitter, in this case the Chief Financial Officer (CFO). Knowing the context of the transmission – that the CFO was e-mailing time-critical information to a validated third party – the transmission was completed without delay.

This additional contextual information, beyond the simple fact that the content violated normal data security policies, offers a clear case of content that an organization would not want to block because of the potential for hampering business operations.

Understanding the varying degrees of control required by the organization, at various points on the network, can drive the level of control needed to adequately protect data.

As a content-centric data security technology, Proventia Content Analyzer offers a diverse and extensible set of search criteria without requiring additional hardware or software purchases for owners of Proventia Network IPS appliances.

Protecting Content with Proventia Content Analyzer Technology

Proventia Content Analyzer is included as part of the latest firmware code base in Proventia Network IPS G/GX appliances. It leverages existing protocol analysis module (PAM) technology included in many IBM Proventia products to identify sensitive data in various protocol and content delivery types.

Incorporating more than 20 different inspection technologies, the PAM works in conjunction with Proventia Content Analyzer to inspect unencrypted data across multiple content types and file formats (including zip and gzip).

Updates to the PAM and Proventia Content Analyzer are made through automatic content updates, ensuring that the most current research intelligence from the IBM Internet Security Systems™ X-Force® research and development team is available at all times.

Proventia Content Analyzer can inspect and identify data in a number of Application Layer Protocols:

- *AIM™ (OSCAR, TOC/TOC2)*
- *Microsoft® Messenger (MSNP9, MSNCL, MSNP10)*
- *Yahoo!® Messenger (YSMG)*
- *IRC*
- *HTTP*
- *FTP*
- *SMB*
- *SMTP*
- *IMAP*
- *POP3*

Along with deep packet inspection of the data transmitted via these protocols, Proventia Content Analyzer can also inspect attachments for content that violates one of the content enforcement signatures. The content types and markup languages that Proventia Content Analyzer can inspect include:

- *Microsoft Office documents*
- *Adobe PDF*
- *Rich text format (RTF)*
- *Text*
- *XML*
- *HTML*
- *GZIP (compressed)*
- *ZIP (compressed)*

With its eight defined signatures, and the ability to define an additional eight custom signatures using the IBM Internet Security Systems™ library of Deterministic Finite Automata (DFA) regular expressions, Proventia Content Analyzer offers a flexible and convenient way to gain data awareness in organizations of all sizes.

Deployment Considerations

As with nearly all new technologies, the question of deployment location is often asked. Current Proventia Network IPS G/GX customers have answered this question since Proventia Content Analyzer is part of their existing IPS infrastructure.

For new customers, Proventia Content Analyzer technology is available as soon as the Proventia Network IPS G/GX appliance is online and the Proventia Content Analyzer feature activated. A Proventia Network IPS G/GX can be positioned anywhere in the network; however, its location and the model of Proventia Network IPS ultimately selected will be largely dependent on the specific network architecture.

When deciding on the particular model of Proventia Network IPS and the number of IPS devices necessary, the following minimum variables must be considered:

- *Performance*
- *Latency*
- *Connections per Second*
- *Media Type*
- *Port Density*
- *High Availability*
- *Fail-Over*
- *Asymmetric/Symmetric routing*

In addition to the performance and feature requirements listed above, there are other requirements to consider when determining how many IPS devices are required and where they should be installed. The rule of thumb for the placement of a security device once the network requirements are met, is anywhere in the network that there is a change in security policy or between zones that are not under corporate control, including the perimeter, the VPN gateway, DMZ, secure enclaves, b2b connections, the data center and between locations of recent mergers and acquisitions, just to name a few. The typical deployment scenario for most deployed IPS/Intrusion Detection System (IDS) devices is at the perimeter in either inline or passive mode. The perimeter is a recommended candidate for enabling Proventia Content Analyzer as it provides immediate visibility into unencrypted content leaving corporate control.

Getting Started – A Primer on Proventia Content Analyzer Functionality

The following section provides a general overview of the deployment and configuration of Proventia Content Analyzer, taking advantage of its pre-defined signatures. The configuration of custom signatures is an advanced activity and is explained in further detail in the current User Guide and/or Online Help distributed with the Proventia Network IPS.

Because Proventia Content Analyzer is an advanced feature of Proventia Network IPS and without proper configuration and turning could disrupt network flows. The feature is turned-off by default until manually configured by a network administrator. The types of content that Proventia Content Analyzer can detect are common on most networks and depending on the location of the deployed Proventia Network IPS appliance, once the feature is activated, it is likely that there will be a high volume of events triggered.

For example, the defined signatures for names, dates and monetary amounts signatures will detect content commonly found on e-commerce sites. In a large organization that allows casual Web browsing, the event queue can quickly fill up as transactions are made and innocuous data content is transmitted. Most organizations will not want to turn on all of the defined signatures out of the box because of the potential flood of events.

Determining a Baseline Configuration

The configuration of Proventia Content Analyzer should be approached strategically, with a goal of creating a content-aware baseline policy first and then gradually adjusting the policy based on the resulting number and types of events and blocked content.

If needs later dictate that all of the defined content signatures be activated, IBM ISS recommends that events be sorted by event type, and then using the resulting data, the parameters further adjusted until the desired level of event alerts and blocking activities are reached.

While making the baseline policy and resulting adjustments, it is important to understand the value, or sensitivity level of data, which will be different in every organization. In organizations that have to adhere to PCI-DSS, the requirements will drive the resulting configuration in determining what content types should receive which type of action.

The following Proventia Content Analyzer baseline policy flowchart and Use Case can provide guidance on building an initial content aware policy.

Proventia Content Analyzer Signature Decision Process

Using the following flowchart, administrators can begin to create a baseline policy based on their organization's particular needs. Once this process is complete, the Use Case in the next section can provide guidance on deployment and configuration.

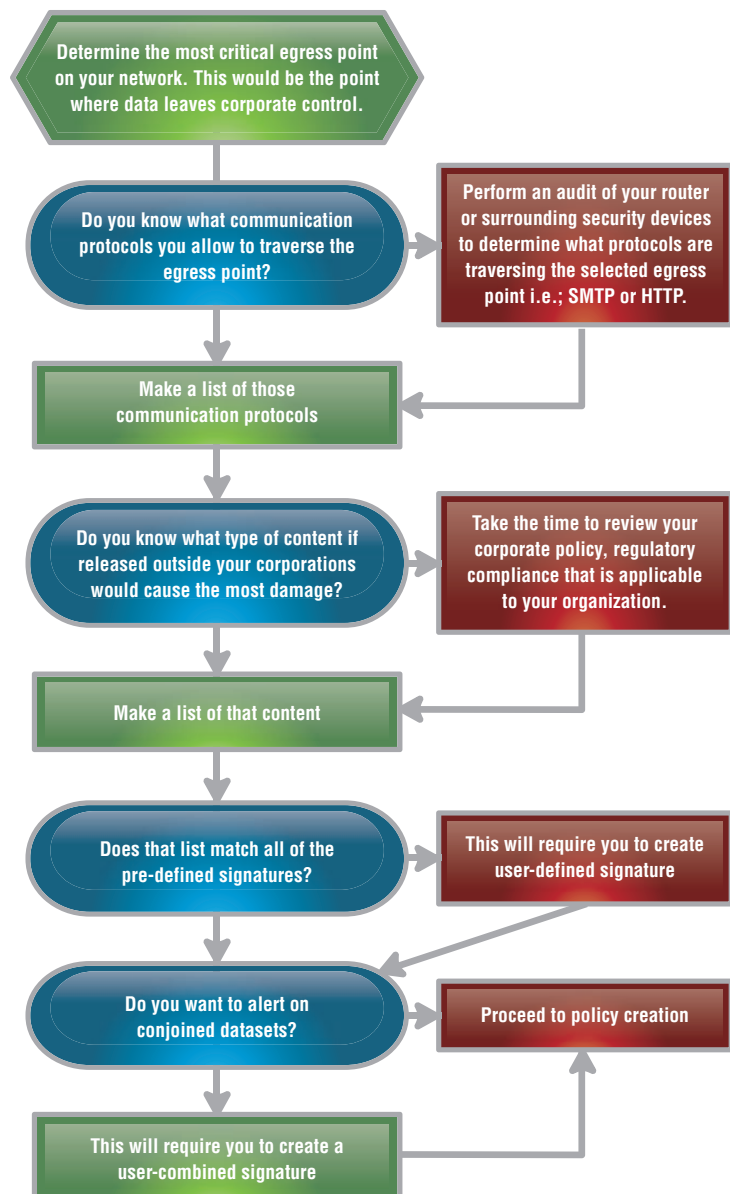


Figure 1: Proventia Content Analyzer Baseline Policy Flowchart

Use Case Deployment Scenario

This Use Case scenario explores the identification, discovery and remediation process an average organization might undertake to help identify and/or block the inappropriate release of sensitive information to third parties using Proventia Content Analyzer, and briefly explains the necessary configuration of Proventia Content Analyzer to address the findings from the discovery process.

Scenario: Retail Company X has determined the need for insight into confidential data that might be leaving their organization unencrypted. Retail X accepts credit card payments, and is therefore responsible for complying with PCI-DSS requirements and undergoing annual compliance audits. In addition to regulatory compliance, Retail X is concerned about confidential project names and details being posted to a blog and more importantly, the accidental or malicious disclosure of client data. After utilizing the Proventia Content Analyzer baseline policy flowchart, they were able to determine the following information:

Deployment: After reviewing all data egress points, Retail X decided to activate the Proventia Content Analyzer feature included on their Proventia Network IPS GX5108 deployed at the perimeter of the network.

Communication Protocol Audit: After reviewing their perimeter router access control lists and firewall rules, Retail X determined that they do not allow inbound connections from the Internet into their network. However, they do allow the following communication protocols to make outbound connections to the Internet: HTTP, SMTP, DNS and Yahoo! Messenger.

Content types: After reviewing their corporate policy and regulatory compliance, and after interviews with the director of product management and human resources, Retail X determined that the following types of information represent potential regulatory and policy violations if disclosed externally either separately or in combination with each other:

- *credit card numbers*
- *names*
- *date*
- *social security numbers*
- *project code name “viner”*

Based on the data above, Retail X decided to leverage Proventia Content Analyzer to help enforce their internal data security policies. The following section outlines the basic configuration steps required to configure Proventia Content Analyzer to begin protecting the organization from unauthorized PII and other sensitive disclosure, as one part of their data security strategy.

Configuring Proventia Content Analyzer

Retail X determined that the types of content being disclosed outside of the organization's network boundaries are supported as configurable signatures within Proventia Content Analyzer as either preconfigured or user-configured signature categories. The signature categories can be broken down into the following:

- *Pre-Defined Signatures: credit card numbers, names, date, social security numbers*
- *User-Defined Signatures: Project code name "viner"*
- *User-Combined Signatures: Credit card number + name + date*

Activating Detection Signatures

Depending on the type of sensitive information being disclosed or based on the organization's data security policies, different actions may be warranted for the disclosure of one type of PII content (e.g., name), while another action may be warranted for that same type of PII in conjunction with another (e.g., name + credit card number). Because of this, a combination of single and multi-expression signature actions should be used to generate alerts as needed or to block the content completely.

Of the eight defined signatures resident in Proventia Content Analyzer by default, the following signatures should be turned on to address some of Retail X's content types:

- *Content_Analyzer_Credit_Card_Num*
- *Content_Analyzer_Social_Security_Num*
- *Content_Analyzer_Date*
- *Content_Analyzer_Person_Name*

Using the additional eight user-customizable signature slots, the user would configure and activate the following signatures:

In the single User Defined signature section, they would turn on the following:

- *Content_Analyzer_User_Defined_0*

This custom signature looks for the customer content “viner”.

In the combined User Defined signature section, they would turn on the following:

- *Content_Analyzer_User_Combined_0*

This custom signature would search for all three of the following types of content within the same transmission:

- *Credit Card Number*
- *Name*
- *Date*

These configuration steps are performed through simple input driven fields, thus reducing the need for the user to understand how to build DFA regular expressions.

Configuring Protocols

By default, all of the protocols that Proventia Content Analyzer can inspect are turned to “on.” Retail X determined they only needed certain protocols inspected, so they configured Proventia Content Analyzer to only inspect the following protocols:

- *HTTP*
- *SMTP*
- *Yahoo!IM*

Configure Occurrence Thresholds

By default, each defined signature has a set occurrence threshold of either 10 or 100 depending on the particular signature. When this threshold is reached, the configured action will occur (i.e., alert generated and/or content blocked). These thresholds may be high or low depending on an organization's particular needs. In Retail X's case, they decided to modify the thresholds to the following values:

- *pam.ca.credit_card_num.minmatch - threshold value = 1*
- *pam.ca.social_security_num.minmatch - threshold value = 1*
- *pam.ca.date.minmatch - threshold value = 1*
- *pam.ca.person_name.minmatch - threshold value = 1*

With these threshold settings, even one event matching the configured signature will trigger an alert.

Additional Use Cases

Use cases will vary depending on regulatory compliance, the particular organization's policies and the value placed on network data. Again, it is recommended that organizations approach the use of the Proventia Content Analyzer feature with a strategy in mind using the Proventia Content Analyzer baseline policy flowchart as a starting point. The following sections offer additional examples of areas of governance in which Proventia Content Analyzer can provide insight into confidential data.

Regulatory Compliance

The Payment Card Industry Data Security Standard addresses the mandatory requirements that a business must have in place in order to secure cardholder and sensitive authentication data. PCI-DSS, like most regulatory compliance initiatives, is very thorough and includes a vast amount of technologies deemed necessary to safeguard customer data. In the case of network intrusion detection and prevention, PCI-DSS v1.1 offers guidance in requirements 10 through 12. However, with Proventia Content Analyzer, customers now have the capability of addressing requirement 4, which is, “Encrypt transmission of cardholder data across open, public networks.”

Specifically, requirement 4.2 states that cardholder data should never be transmitted unencrypted via e-mail. Since Proventia Content Analyzer has the capability of alerting on or blocking single and multi-criteria content (e.g., name + credit card number) across a number of today’s popular protocols, organizations now have the capability to monitor compliance and compliance violations, pursuant to Requirement 4.2 with Proventia Network IPS. This offers the ability to receive immediate notification of PCI-DSS and other regulatory policy violations.

This function is also useful as a verification tool that the network is configured as expected and data are not being transmitted unencrypted against the organization’s wishes. This offers administrators the capability to remediate those issues before an audit or public incident, and it also provides insight into the accidental or intentional unencrypted transmission of sensitive data.

As part of the Proventia Network IPS, Proventia Content Analyzer can also be utilized for network segmentation that isolates the cardholder data environment from the rest of the network by monitoring that border for PCI related data. Network segmentation of cardholder data may help your organization comply with specific PCI-DSS requirements and may help reduce the overall PCI-DSS scope.

Corporate Policy:

Corporate policies regarding data privacy and security vary greatly depending on the nature of the organization and its particular governing structures. However, the accidental or intentional disclosure of employee data, client data, project names and e-mail destined for nefarious destinations are common examples of information that must be protected or brought to attention when transmitted outside the organizational infrastructure. The protection of this information is commonly documented in the form of an Acceptable Usage Policy (AUP) and corporate confidentiality guidelines. The Proventia Content Analyzer feature helps organizations alert and identify/enforce AUP violations and corporate confidentiality guidelines.

Conclusion

As the Internet becomes omnipresent in organizations through the use of online collaboration tools such as wiki's, blogs, and instant messaging, VoIP teleconferencing, etc., the need for visibility into the data leaving the network boundaries will continue to garner attention, while the ability to provide awareness into that content will continue to challenge network administrators. In response, organizations should develop a comprehensive strategy to address these data security challenges.

IBM Proventia Content Analyzer is a unique technology that allows organizations to utilize their existing IBM Proventia Network IPS infrastructure to inspect and block unwanted PII and sensitive content disclosure across multiple protocols. With the proper tuning and security strategy, IBM Proventia Content Analyzer is an effective tool to help safeguard sensitive information that will complement a holistic data security solution.

About the Author

John W. Pirc. As Senior Product Manager for IBM ISS, John is responsible for the direction of the IBM Proventia Network IPS G/GX appliances that comprise IBM's extensive security product portfolio. In addition to providing strategic direction for the Proventia G/GX product line, John works very closely with the X-Force team to ensure that critical security content is integrated into the product line. Prior to IBM, John worked as a Product Manager for Cisco's IPS product line and the U.S. Intelligence Community. John has more than 10 years of security experience in security research, security IV&V testing, forensics, and architecting/deploying enterprise wide security solutions for both public and private organizations worldwide. In addition to a BBA in Information Systems from the University of Texas, John also holds the NSA Information Assurance Methodology and Certified Ethical Hacker certifications.

About IBM ISS

IBM ISS is the trusted security expert to global enterprises and world governments, providing products and services that protect against Internet threats. An established world leader in security since 1994, IBM ISS delivers proven cost efficiencies and reduces regulatory and business risk across the enterprise. IBM ISS products and services are based on the proactive security intelligence conducted by the X-Force team – a world authority in vulnerability and threat research. For more information about IBM ISS, please contact your IBM representative or IBM Business Partner. You may also call 1 800 776-2362 or visit **ibm.com**/services/us/iss.



© Copyright IBM Corporation 2008.

IBM Global Services
Route 100
Somers, NY 10589
U.S.A.

Produced in the United States of America

02-08

All Rights Reserved

IBM and the IBM logo are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both.

Proventia and SiteProtector are trademarks or registered trademarks of Internet Security Systems, Inc. in the United States, other countries, or both. Internet Security Systems, Inc. is a wholly-owned subsidiary of International Business Machines Corporation.

Other company, product and service names may be trademarks or service marks of others.

References in this publication to IBM products or services do not imply that IBM intends to make them available in all countries in which IBM operates.