GIAC
CERTIFICATIONS

Global Information Assurance Certification Paper

Copyright SANS Institute
Author Retains Full Rights

This paper is taken from the GIAC directory of certified professionals. Reposting is not permited without express written permission.

Interested in learning more?

Check out the list of upcoming events offering
"Security Essentials: Network, Endpoint, and Cloud (Security 401)"
at http://www.giac.org/registration/gsec

Data Loss Prevention

GIAC (GSEC) Gold Certification

Author: Randy Devlin, rdevlin@mastersprogram.sans.edu

Advisor: Stephen Northcutt

Accepted: September 6th, 2015

Abstract

Data Loss Prevention (DLP) continues to be a complex business-centric security initiative for organizations to overcome. The complex nature is mainly attributed to the multiple attack surfaces and limitless exfiltration scenarios. This research is not focused on the network or system compromise. Being compromised is a lower category of criticality than being breached; a compromise is a precursor event to the breach. This research is targeted at the activities surrounding detection and exfiltration. The complexity is not necessarily a technical challenge, rather more so the numerous scenarios that a DLP program must differentiate between "known-good" activity and the commensurate design of the detection settings to catch exfiltration; these are iterative and potentially infinite. Defining "known-good" activity enables the business to continue to operate. Designing detection settings for the exfiltration requires a DLP engineer to think as the attacker and obtain visibility within the multiple attack surfaces that may be leveraged in staging the exfiltration. The research and scenarios in this paper demonstrate how the network detections are capable of being bypassed by well-defined exfiltration techniques. The research leverages commercially available platforms. The goal of the findings is intended to continue to enhance and advance the organizational DLP programs chartered to defend against the threat of losing sensitive data.

© 2016 The SANS Institute Author retains full rights.

1. Introduction

Data Loss Prevention is a security challenge that is quite unique when compared to traditional enterprise security layers of protection. Traditional enterprise security protection layers examples include Network Intrusion Detection Programs and Endpoint Protection Programs. DLP Program attributes that differ from traditional security protections are the inclusion of several business stakeholders to the DLP program, business liability for data inspection, and program lifecycle management (Kumaresan, 2014, page 2). These attributes require coordination, policy, and detail oriented analysis, leading to the increased program complexity.

Detecting and preventing the loss of data can prevent brand damage, competitive disadvantage, and/or legal proceedings. The DLP program is the mechanism by which an organization identifies their most sensitive data, where the data is authorized to be stored or processed, who or what application(s) should have access to the data, and how to protect from the loss of the sensitive data. The sections below provide a framework for defining components of the DLP program (Sullivan, 2008, page 3).

The next sections will outline the components that comprise DLP strategy and determine their level of inclusion in the DLP program. These components include data types, data classification, and threat actors to data (Kumaresan, 2014, page 2). During the lifecycle management of the DLP program, these components must all be periodically defined, re-evaluated and evolved.

1.1 Data Types

Data is stored in two different manners, structured and unstructured. The type of process using the data dictates the type of data. Functional examples for the usage of structured and unstructured data are demonstrated by Nemschoff (2014).

A classic example of structured data is a database process that stores and indexes binary digits in a structured manner, allowing for reference and repeatable associations or recall. The inputs and outputs of each process are repeatable and predictable, which are characteristics of

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

structured data types. The predictable usage is finite; therefore, DLP inspection logic around structured data is finite.

Unstructured data types may include document processing, emails, short message service (SMS) notifications, audio/visual interactions, or imagery. DLP faces a challenge with these data types as they provide random and infinite processing. Unlike structured data types, unstructured data types do not provide repeatability or predictability. For example, document processing applications that allow users to manipulate and save individual documents are considered unstructured data. However, determining what makes the content or context of the document sensitive becomes the challenge, so this is not a finite process.

1.2 Data Defining and Classification

Defining data types serves two purposes within the DLP Program. First, after defining the data type, the organization understands the data use and knows the limited locations where the data might exist. Second, defining the data type enables the organization to determine a method for classifying the data type. Defining the data type will determine whether the data is structured, unstructured, or possibly both.

In classification, the organization defines the attributes of the data to ensure that the detection technologies can identify the data and handle them as policy dictates. Sensitive data classification types help the DLP program determine the detection capabilities required to alert the organization and prevent violations from occurring with the data.

Classification structure is extremely important for a successful DLP program. Data is constantly changing location, user, and type. Constant classification re-evaluation and potentially changing inspection locations, classification structure should be capped at no more than five classifications or policies. Standard built-in policies can reduce time to value and are less complex to implement (Ernst & Young, 2011, page 18). These built-in policies are centered on data types with traditional use cases. Non-standard use cases, such as looking for intellectual property, will add complexity.

© 2016 The SANS Institute Author retains full rights.

1.3 DLP Threat Actors

DLP was originally designed to alert organizations to the unintended misuse of internal data by an organizational employee, identifying broken business processes during the discovery (Kanagasingham, 2008, page 6).

The design targeted a non-malicious insider threat. This threat is common when an employee with appropriate access to data inadvertently misuses the data with the right intention, breaking data governance policy or process. As a simple example, Employee X copies classified data to a personally owned external media device to complete the task at home, before returning the data. Standard data governance policy does not allow classified data to be copied to non-classified assets, nor does it allow data to be copied or manipulated on non-corporate systems.

As time progressed and technology advanced, the use cases for DLP evolved. The system and network compromises that lead to data breaches by malicious users increased significantly. This increase prompted a reevaluation of the threat actors interested in accessing, stealing, or destroying data. This reevaluation, in turn, defined two additional threat actors, malicious insider and malicious outsider.

A malicious insider threat is an employee intending to break the data governance policy. The reasons for this behavior may vary; the employee could be resigning, the employee could be aware of an impending demotion, or the employee might have been convinced by a third party to steal the data.

A malicious outsider threat actor is not affiliated with the organization. The threats from outsiders can come from competitors, "enemies," or outsiders who intend to capitalize on the resale of an organization's data. Outsider DLP threats can attempt to coerce companies to cease current actions or to influence decision making. Examples of this include the attempt to prevent Sony Pictures from releasing a movie that was disapproved by certain groups (Peterson, 2014), or Ashley Madison where the attackers wanted the site to cease operations. These styles of attacks are characterized as hacktivist vigilantism (Basu, 2015). Classifying outsiders as malicious can be subjective, based on the beliefs of those involved. Either way, the legalities of privacy remain a factor.

© 2016 The SANS Institute Author retains full rights.

2. Technology Deployment

The determinations made for the components that comprise DLP strategy directly impact the deployment of the technology. DLP technology deployment centers around three key elements: inspection visibility, inspection capability, and detection response

2.1 Inspection Visibility

Detection is predicated on visibility and content interrogation and/or understanding. Without achieving both DLP inspection is impossible.

Visibility is the placement of the DLP solution in a manner that the platform has full access to where the data is to be inspected. To understand, a DLP engineer first defines where the sensitive data resides and what ability to inspect exists. Three inspection approaches exist, scanning data at rest, scanning data in motion on the network, or scanning data on endpoints.

Scanning data at rest places a component of a DLP platform close to the sensitive data and scans searching for sensitive content. These scanning platforms are generally directed towards network share locations, long term storage, database backups, or archive storage locations. Two items are of significant importance with data at rest scanning:

1. Negative network impacts exist the further away the DLP component is scanning from. The data at rest scanning components can be as "noisy" or more so than vulnerability scans.
2. Understanding records management and backup strategies are key, as by default a data at rest scanner opens files or modifies "last modified" dates and, therefore, can create contention between backup solutions and DLP scanning periodicity.

Scanning data in motion on the network is a DLP inspection capability that allows the DLP platform component to parse network based protocols capable of data transfer in the payload. To keep up with the speed of the network, some of these components are developed to focus on limited amount of protocols; for example DLP for email would focus on simple mail transfer protocol (SMTP), post office protocol version 3 (POP3), Lotus email, and internet message access protocol (IMAP). Other network-based DLP components focus solely on the inspection of

© 2016 The SANS Institute Author retains full rights.

high-risk protocols capable of data transfer, such as instant messaging (IM) and file transfer protocol (FTP). The components of network DLP can be architected inline to be included in the data flows (content flows in, inspected, then forwards out a separate interface) or can inspect out of band via tap capability.

A few considerations must be taken in order to properly architect network based DLP. The placement of DLP inspection platforms can be designed for inline or in tap mode. Inspection is conducted the same with either placement, but the response capabilities significantly differ between inline or tap.

The inline mode will require two interfaces, one on each side of the connection. In order for communication to occur between two systems, the session must go through the DLP appliance. In going through the appliance, the session can be evaluated via policies for unauthorized content. If unauthorized content exists, the DLP appliance can take all of the response techniques covered later in section 2.7.

Tap mode requires a single interface and is forwarded a copy of the data for inspection. The fact the session does not traverse the appliance, response mechanisms such as dropping the connection are not possible within the DLP appliance. Two main response mechanisms exist; for transmission control protocol (TCP) connections, a reset may be configured to drop the connection. The second and least capable response is sending a flag to the tapping device that a violation has occurred. This leads to a further configuration, if available, on the tapping device to handle the violation.

In considering placement, a second, more challenging situation exists. Canadian based company projected by the end of 2016 that the continent's traffic ratio of encrypted to unencrypted connects will reach 90% to 10% respectively. This trend can be further confirmed with the migration of video streaming leader, Netflix, moving to an encrypted streaming model (Hackett, 2015). The encrypted traffic requires a second step to be accomplished for DLP appliances to effectively inspect; the payload of these connections must be decrypted.

Network traffic decryption has remained challenging on two significant fronts, capability, and privacy (VanAntwerp, 2011, page 21). Decrypting traffic for inspection in real time requires

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

dedicated hardware specifically designed for decryption and re-encryption. This process can add two to three times the existing latency due to the addition processing.

The challenges with privacy exist around the type of data that is being decrypted. Network users may have an expectation of privacy when banking online or reviewing sensitive healthcare information. Canada and the European Union (EU) have "heavy" data regulation and enforcement policies in place further protecting citizens (DLA Piper, 2015) For these privacy reasons, legal departments must be involved early and often in the discussions of DLP platform architecture.

The third component of detection visibility is scanning data on endpoints. Fundamentally, this component is far different than the previous two. This solution is software based on the endpoint and has full visibility to the system the DLP software is installed upon. For visibility, the endpoint DLP agent should have access to the network stack (so similar to network DLP) and also file level access to the endpoint.

Detection of sensitive data on endpoints may rely upon an understanding of both structured and unstructured data. Sensitive data in text documents and the various methods of data and document manipulation is the focus of the unstructured detections. Copying, pasting, saving as different file types and encrypting are some of the manipulations that the DLP agent would need to evaluate during inspections.

Structured sensitive data inspection on endpoints revolves around the applications that users utilize. The increase in the use of web applications has led to an increased need for these inspection methods. The data is structured due to the process repeatability, fixed values and distributed nature of sharing. The constant proliferation of mobile device users further increases the risk to sensitive structured data leakage, as application owners release mobile applications, on non-corporate devices, allowing access to the data.

2.2 Inspection Capabilities

Once detection visibility is achieved, inspection capabilities follow. Being placed properly on the network or endpoint is different than the ability to actually detect the movement or handling

© 2016 The SANS Institute Author retains full rights.

of the data. A helpful analogy is a person being able to see (inspection visibility) and knowing what the person is looking for (inspection capability).

Inspection capability centers around two main categories for DLP, context inspection, and content inspection. Data inspected by its context is characterized by the location, application usage, and/or users. Context inspections can be considered as metadata of the actual binary data. Metadata is attribution or characterization of or about another item. The sensitive location of the data is NOT actually the data, but ensuring the data is not transferred from the sensitive location is the responsibility of the context inspection. Another example is a document that has a meta tag associated with the document so a context inspection will handle the document in accordance with the configured policy response.

Data inspected by content is characterized by tags or markings, exact data matching (EDM), indexed document matching (IDM), and data string matching. Content inspection can be further differentiated between low complex and high complex. Low complex content is an inspection technique that leverages tags, keywords, regular expressions or other finite simple inspection criteria.

Low complex content inspection has fewer performance impacts and is unaware of anomalous content activity. Tagging or marking content is a method to quickly identify the sensitive content from non-sensitive content within the stream. The tagging or marking of the data itself and is different than a meta tag used during context inspection. The challenge of tagging or marking is the preservation of the tags/marks throughout the data lifecycle. The moment processing of content removes a tag or mark, the DLP content inspection capability is lost unless the DLP inspection platform is designed with persistence of the data during processing; not all vendors are persistent.

EDM leverages the understanding that the content is indexed in a database or other tabular or relational format. EDM is a structured data content inspection method. IDM is an unstructured data content inspection method that focuses on indexing documents not in a database or other organized, repeatable method. Where IDM is different than EDM is that the DLP platform must have some or all of the sensitive components of the document to match on because the source will be unknown as in EDM.

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

Data string matching relies upon the DLP policy configuration to understand portions of what the sensitive data is within the content. A common example is the search for credit card numbers. A sample regular expression (regex) is shown in the below figure depicted (Finding or Verifying Credit Card Numbers, 2013):

```
^(?:4[0-9]{12}(?:[0-9]{3})?          # Visa
 |  5[1-5][0-9]{14}                  # MasterCard
 |  3[47][0-9]{13}                   # American Express
 |  3(?:0[0-5]|[68][0-9])[0-9]{11}   # Diners Club
 |  6(?:011|5[0-9]{2})[0-9]{12}      # Discover
 |  (?:2131|1800|35\d{3})\d{11}      # JCB
)$
```

Figure 1. Regular Expression Data Matching Example

The majority of the credit card issuers have a constant leading digit followed by a series of variable digits with varying delimiters or dashes. Regular expression matching allows the content inspection to utilized patterns and logic to determine the difference between sensitive data and non-sensitive data.

In addition to standard credit card formats, credit card issuers leverage an algorithm known as the Luhn algorithm to mathematically determine valid credit card numbers to issue to customers (Finding or Verifying Credit Card Numbers, 2013). It is the combination of regex detection and Luhn algorithm validation that enables DLP inspection platforms to limit false positives with a set of credit card numbers that simply match the regex pattern.

High complex content inspection is commonly associated with the optical character recognition (OCR) capability. A DLP platform capable of OCR inspection is capable of looking for the low complex content inspection within multiple image formats. An example is sensitive data on an authorized endpoint is opened with Microsoft Word. The user then takes a screen capture as the document is opened and saves the file as a Portable Network Graphics (PNG) format. Low complex content inspection methods are unable to detect the sensitive content within the PNG file, whether detection visibility is positioned on the endpoint or network or both.

© 2016 The SANS Institute Author retains full rights.

OCR inspection capability would leverage the low complex content rules to "visually" detect the sensitive data within the PNG file. The nature of the complex OCR inspection requires robust processing and memory capacity; for this reason, OCR is uncommon for network detection visibility. Network detection in near real time happens at a rate that is too fast for OCR to keep pace.

2.3 Detection Response

Response techniques are actions that can be taken to prevent unauthorized data sharing. Merely alerting of an incident can be quite useless when dealing with data theft; by the time review of the event occurs it is likely the threat actor has completed the intended data breach.

Response techniques vary based upon the desire to inform the threat actor that the actions are monitored and unauthorized. For example, blocking the exfiltration of data likely will alert the threat actor that their attempt to steal data is blocked and may further prevent the type of exfiltration attempt in progress. While this blocking is extremely effective, it sets up two common scenarios:

1. The attacker may attempt alternate measures different from the blocked activity
2. The DLP analyst will be unable to further obtain attribution of the threat actor

To prevent these fallout scenarios, the introduction of non-blocking response techniques exist, data modification, data sanitization, and data mutilation. These response techniques ensure that some form of data is returned to the attacker, but is unusable. The intent is to protect the data while not discouraging the attacker from continuing to attempt exfiltration; the follow-on attempts are valuable in further collecting the tactics, techniques and procedures (TTP) for the attacker for attribution and infusing the collected TTP into additional security platforms as indications of compromise (IOC's) or attack attempts by the same advisory elsewhere in the environment.

Data modification focuses on modifying the data through bit masking. Bit masking is a response technique leveraging AND, NOT and OR operations to modify the original value specifically to generate another value. Bit masking is used in data compression and graphics design, in order to reduce the file size, but the difference is the sending application in data

© 2016 The SANS Institute Author retains full rights.

compression is aware of the bit masking operations so the receiving application can reverse the operations to render the complete initial file.

DLP leverages data modification for the ability to algorithmically modify the binary level of the data so the end result is unreadable or unrenderable to the attacker. The danger in this type of response technique is the ability for the attacker to leverage reversing algorithms to restore the data to the original form. Data modification can also inject predefined garbled data into the binary data constantly rewriting finite bits over and over. This will be more difficult to reverse due to the injections are not predicated on the underlining data or the algorithmic sequence.

Data sanitization is targeting specific content within a larger data set. The specific data, generally the sensitive content at a minimum, is then simply removed and without padding the content. This response technique likely indicates that the attacker has been detected.

Data mutilation, as named, is a response that mal-forms the data. Data mutilation is a combination of data modification and data sanitization.

Taking an action on a DLP event is the ultimate goal of the DLP program. The nature of constant data growth and changing business processes makes achieving response techniques extremely difficult. Response techniques require constant event trigger policy reviews and subsequent responses as changes to the organization occur, thus preventing false positives and work stoppage.

3. Detection versus Exfiltration

Discussed below is a data leakage technique documented in an October 2014 article posted on Dark Reading (Narayan, 2014). The article does not release the identity of the company that experienced the data exfiltration but defines the levels of methods to obfuscate the activities and avoid detection.

The attacker(s), once on the target system, leveraged “chunking” to segment the data first. Once segmented, the smaller pieces were then encrypted. The encryption was an attempt to prevent the organization from determining what the data actually was. Lastly, the encrypted

© 2016 The SANS Institute Author retains full rights.

segments were embedded into a usable video format so inspection isn't alerted to malformed file types.

The attacker was detected post data exfiltration, the key indicator being the video segments were all the same size. Video protocols typically do not transfer file sizes of the exact same during transfer. This breach exemplifies the sophistication and multi-threaded exfiltration approach DLP analysts must plan and test technology deployment scenarios against.

3.1 Data Exfiltration In-Line Detection

Policy based detections are predicated upon the policies being configured within the detection technology. This particular testing scenario involves two clients (Client A and Client B) on opposite sides of the platform where the policy is configured. The protocol for moving the data is simple message block (SMB), a windows protocol for transferring data in an unencrypted manner.

The policy in this scenario is designed to block file transfers when a single social security number (SSN) is present within the stream. The inspection takes place inline, therefore, is capable of the blocking response technique.

The lab diagram and the policy configuration steps are captured in the Figures 2 through 9. The diagram indicates the position in the network where policy inspection occurs.

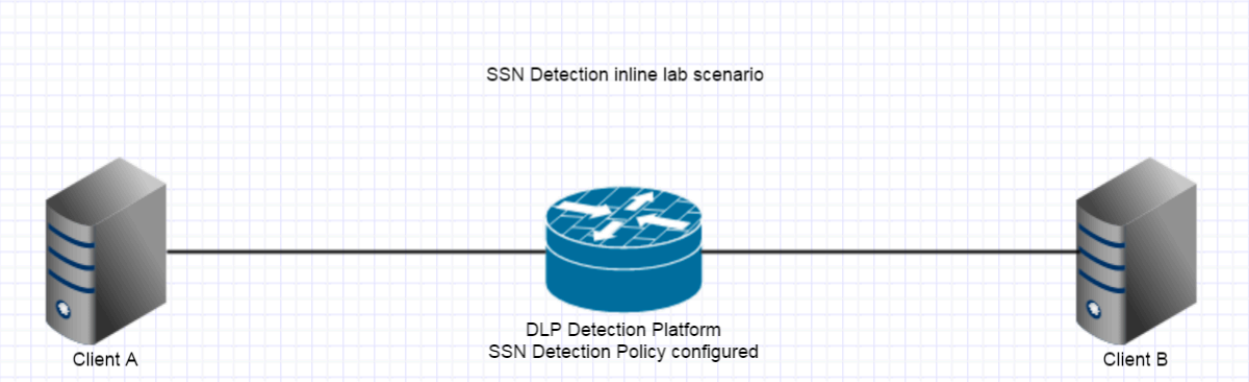

Figure 2. SSN Detection Inline Scenario Diagram

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

Data pattern definition leverages the intellectual property of the security platform to identify SSN's within network data streams. A minimum threshold can be set in the Weight section to respond when the target number is exceeded.

Figure 3. Data Pattern Defined

The data filtering profile inspects specific traffic types for the pattern previously defined. All applications and all file types are inspected bi-directionally. Both the alert and block response techniques are configured, the block will take precedence.

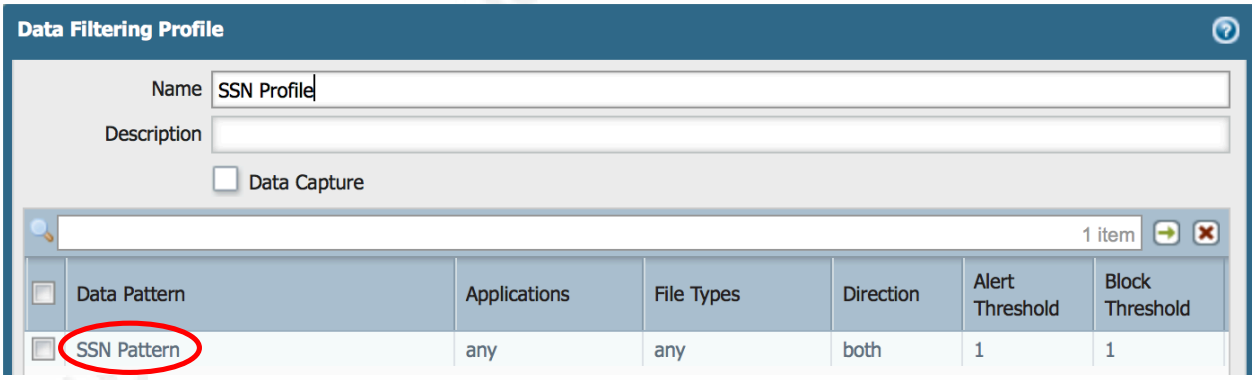

Figure 4. Data Filtering Profile

The data filtering profile is then assigned to a security policy rule. The security policy rule is then applied to specific ingress and egress points of the network for inspection. Data Filtering is set to SSN Profile in this configuration. This will ensure that the security policy rule

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

understands that not only are source and destination important for allowing/denying traffic but that also the content is inspected.

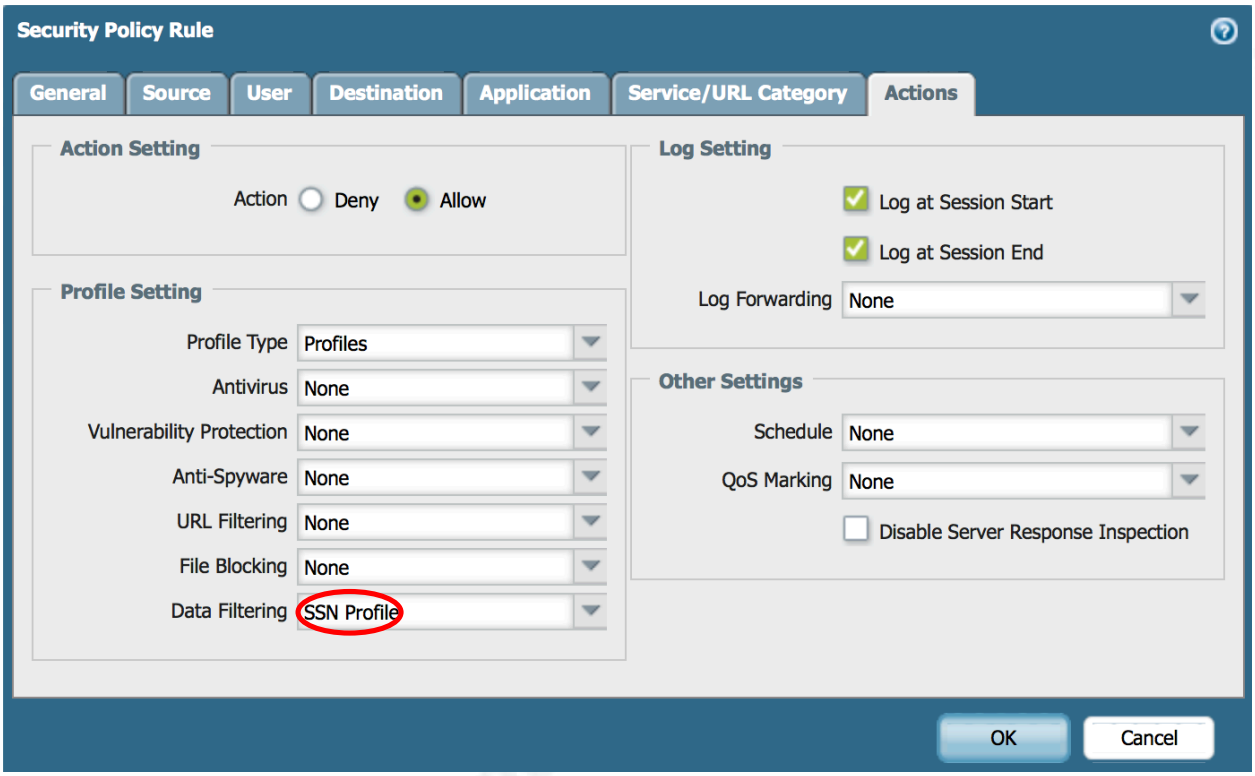

Figure 5. Security Policy With Data Filtering Profile

Below is the file content, current location for the test file, and the desired destination to copy the file. An error is presented to the user that the file is unable to be copied.

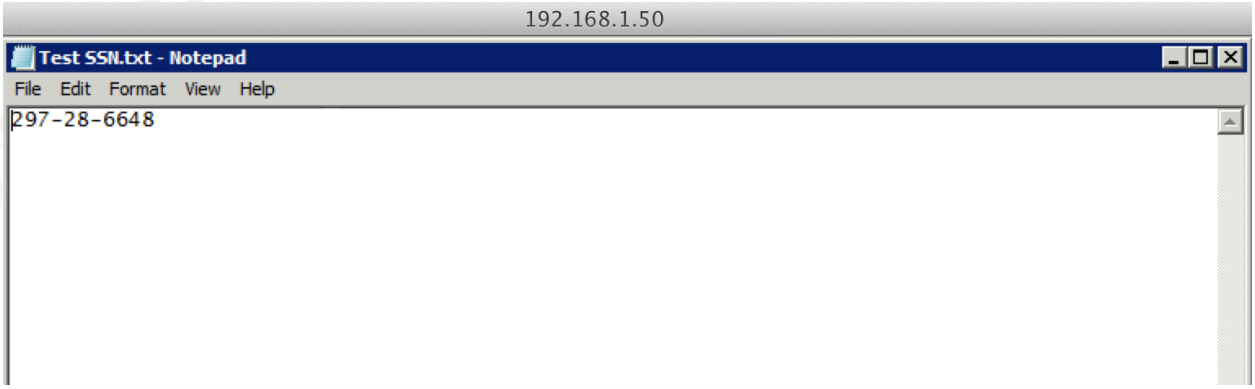

Figure 6. File Content

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

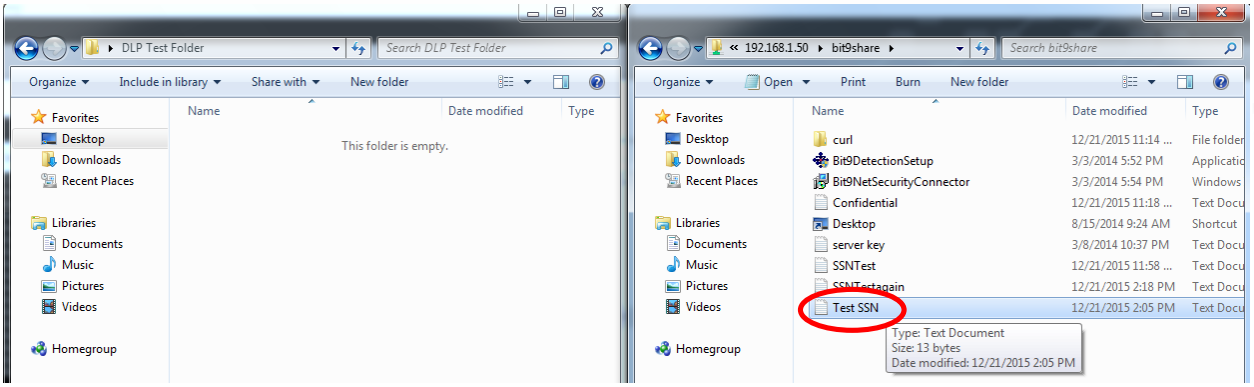

Figure 7. File Location

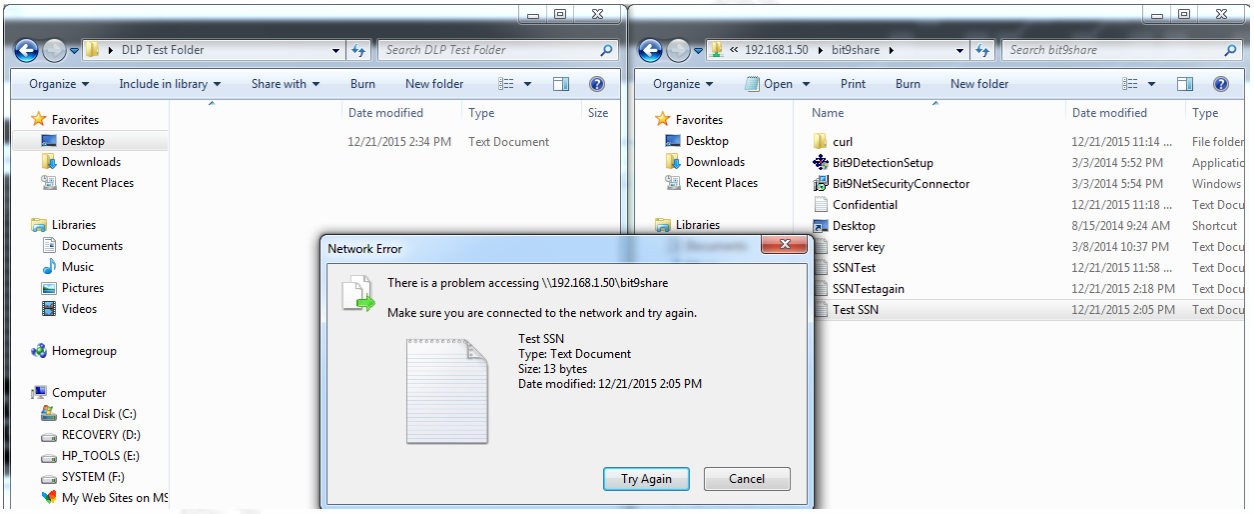

Figure 8. Failed File Copy Attempt

A review of the detailed logs reveal the connection attempt, along with port and protocol (application), the file associated, and the response technique taken.

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

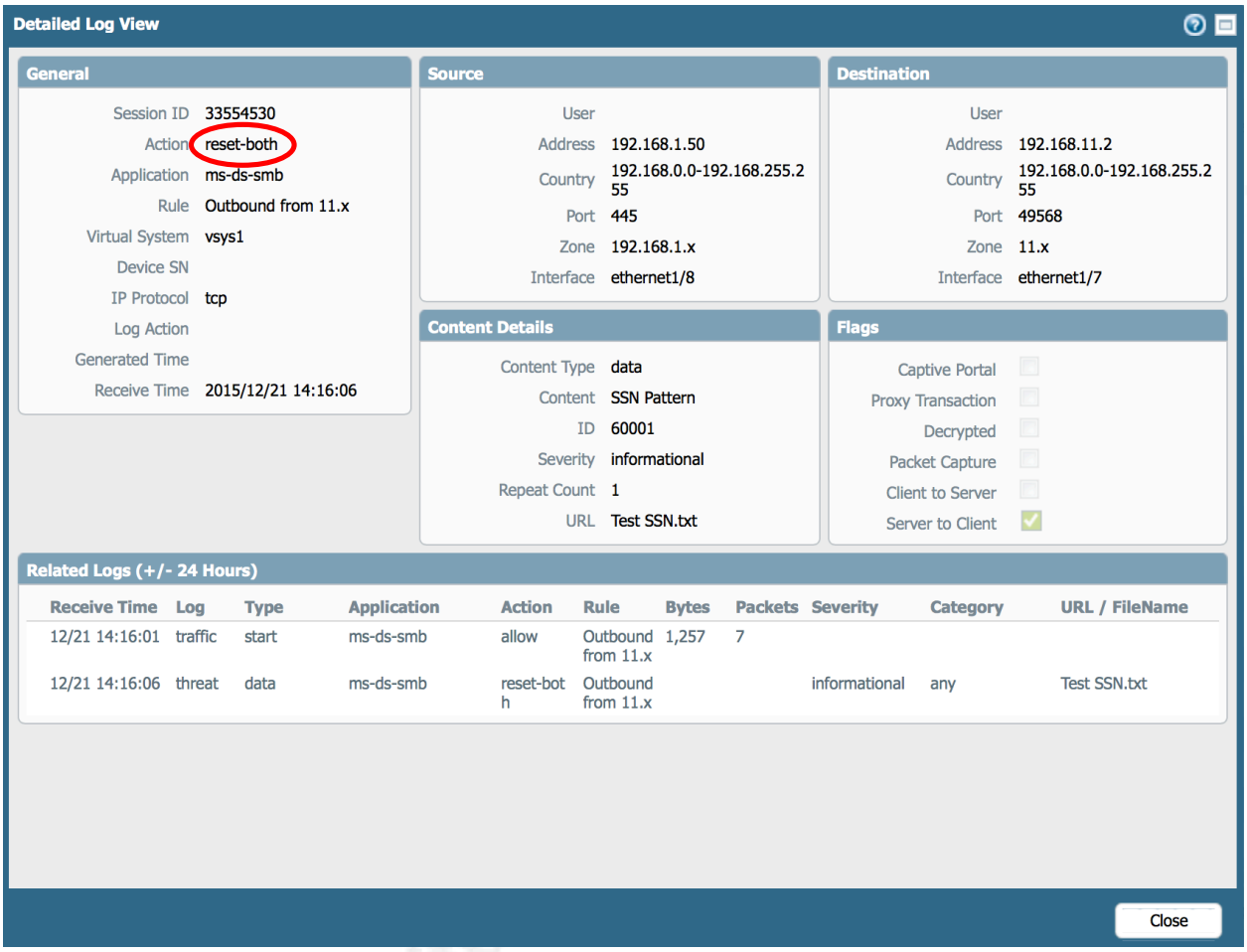

Figure 9. Logging Of Denied Copy Attempt

3.2 Data Exfiltration SPAN/TAP Detection

The network placement for the testing scenario is displayed in Figure 10. The DLP inspection platform is placed off of a session port analyzer network (SPAN) interface. The SPAN port is configured to duplicate the bi-directional traffic seen on the interface Client A is connected to. This duplicate traffic is then forwarded out of the SPAN interface toward the DLP inspection platform; thus providing out of band traffic visibility.

A test word file with SSN data is uploaded from Client A to http://contentiqtest.com. The DLP inspection platform has been configured with Content, Rule, and Policy necessary to detect and alert if the SSN's are seen on the network. Contentiqtest.com is a free site with various test

© 2016 The SANS Institute Author retains full rights.

files for testing DLP policies and rules. Recognize this site does not utilize encryption for data transfer.

The content, rule, and policy configuration is different than the inline lab scenario. This is a different DLP inspection platform than the platform in Figure 2 and allows for additional configuration. The more robust configuration is one of the differences between the platforms. Gartner has deemed the platform in the inline scenario to be DLP Lite, as there are not as many granular inspection capabilities that can be configured for elaborate inspections (Ouellet, 2013).

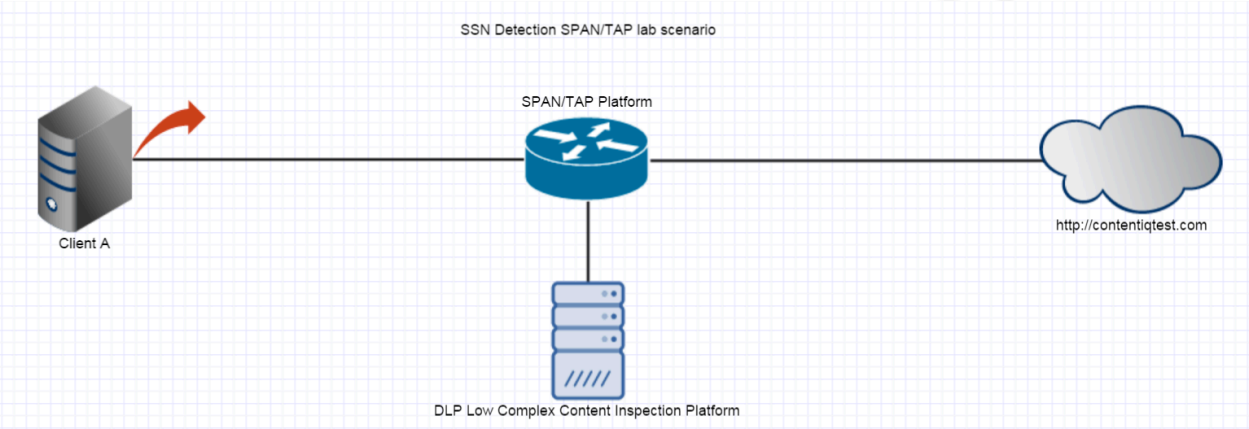

Figure 10. SSN Detection SPAN/TAP Scenario Diagram

Content inspection profile is the configuration aspect necessary for the DLP rule to examine content. The content for the Identity Profile is dependent upon pre-configured expressions to detect for SSN's in various formats; examples include with dashes, without dashes, with spaces, and without spaces. One example of the difference between the full DLP platform and the DLP Lite platform is the capability for content inspection to understand that not all combination of 9 digits (United States SSN) are necessarily SSN's. The number 123456789 is NOT a valid US SSN and therefore, would trigger a false positive SSN detection on the DLP Lite platforms.

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

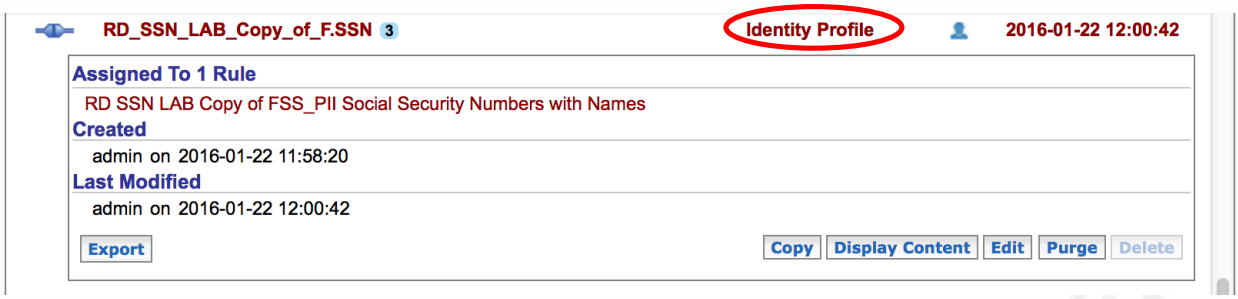

Figure 11. Content Inspection Profile

The identity profile is then associated with the Rule and subsequently the Policy, which if SSN content exists will trigger an Alert notifying the DLP analyst of the incident.

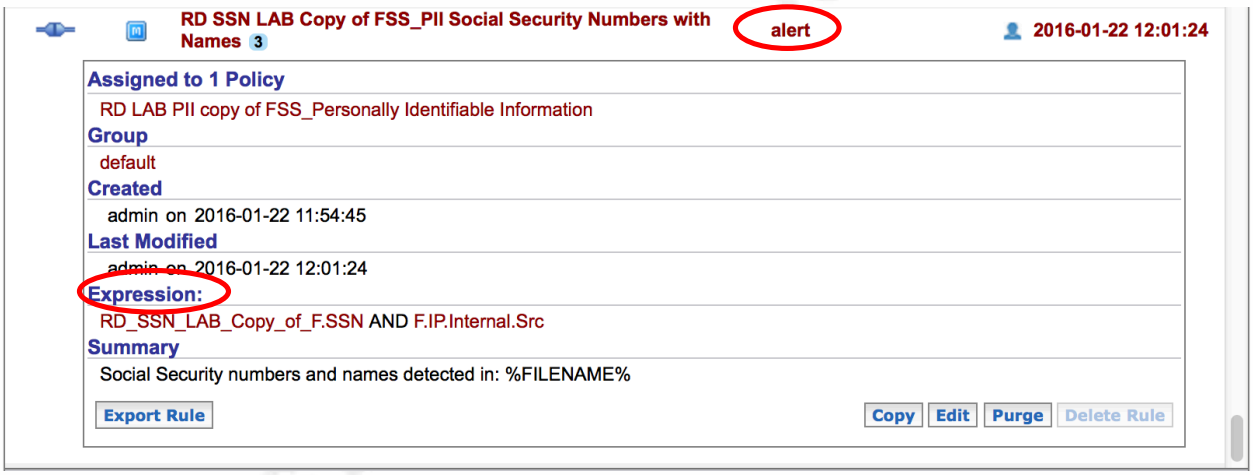

Figure 12. Rule leveraging Content Inspection Profile Expression

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

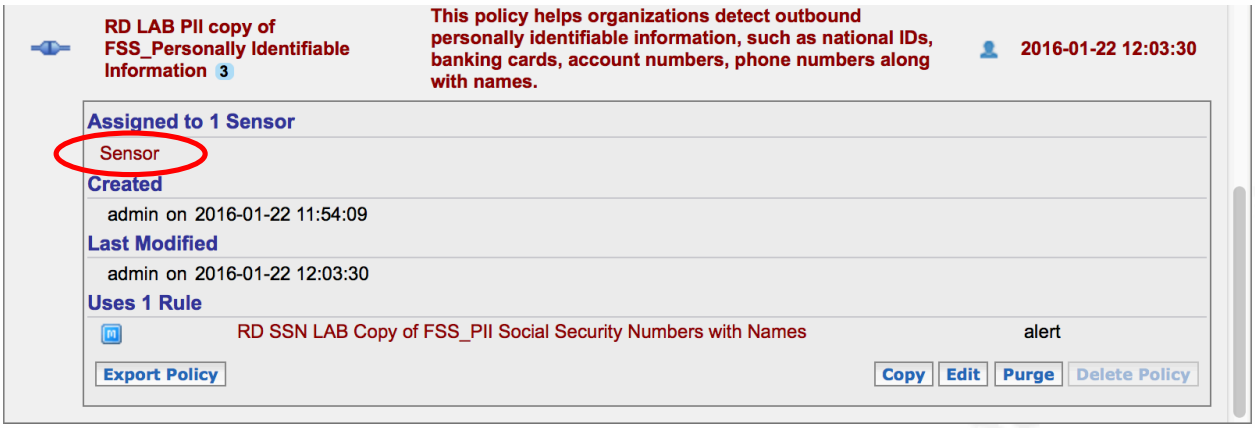

Figure 13. Rule Applied Within Policy

The policy is then assigned to the sensor where the inspection is desired.

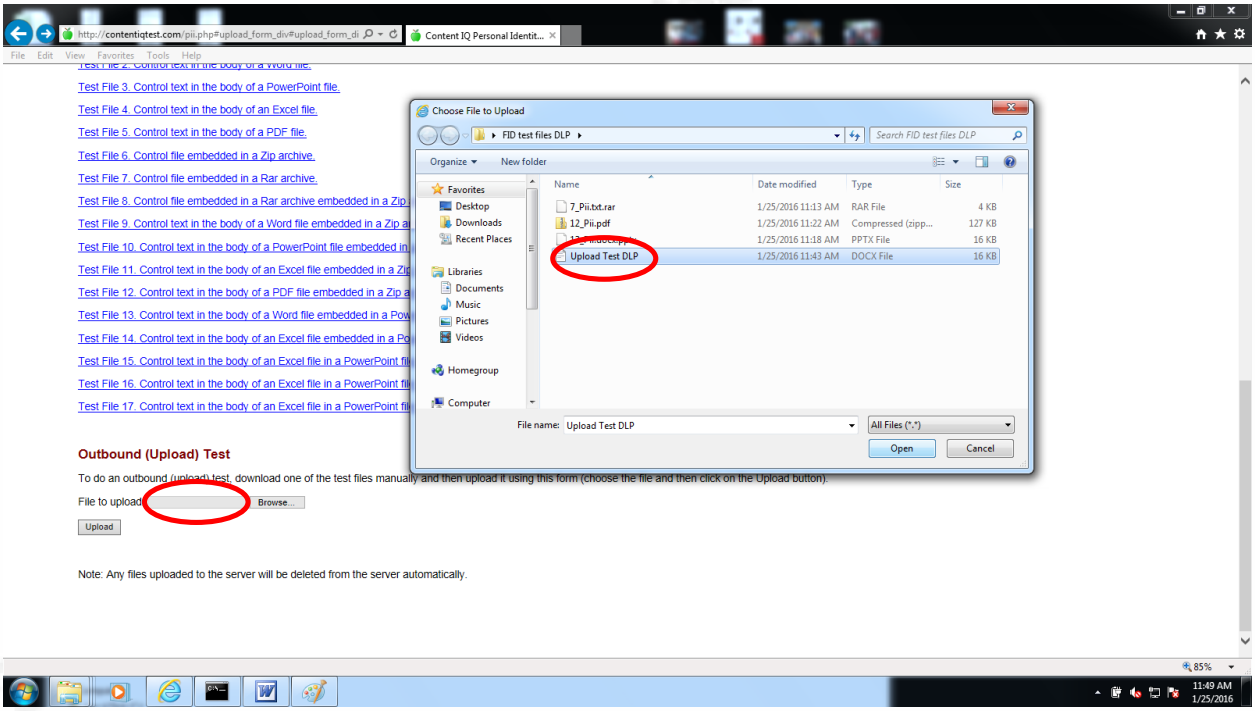

Figure 14. Upload Of SSN Data To ContentIQTest

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

Client A then connects to http://contentiqtest.com and attempts to upload the word file containing the SSN's; the same content found in Figure 21. Below, a high-level incident is generated.

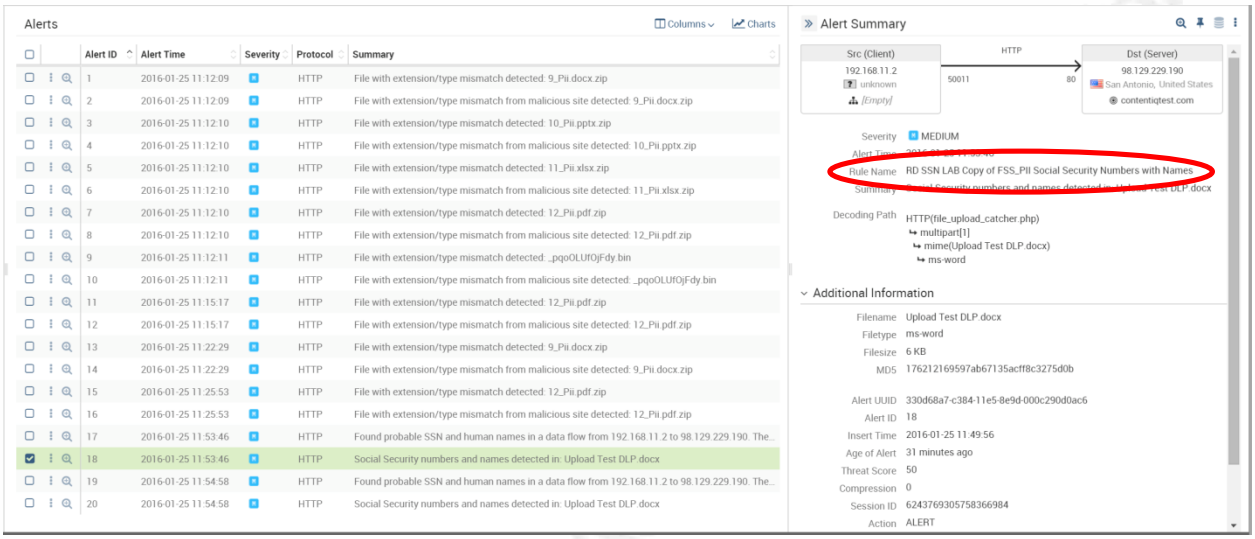

Figure 15. High-Level Alert Response

Further details of the incident are available by selecting from the Alert Summary section. These details further separate the full DLP inspection capability from the DLP Lite solutions in that the content that triggered the incident are displayed for ease of analysis to determine false positive from true positive. The inline scenario sample triggered the incident and blocked the file, but limited forensic evidence is available within the DLP Lite platform to make the determination.

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

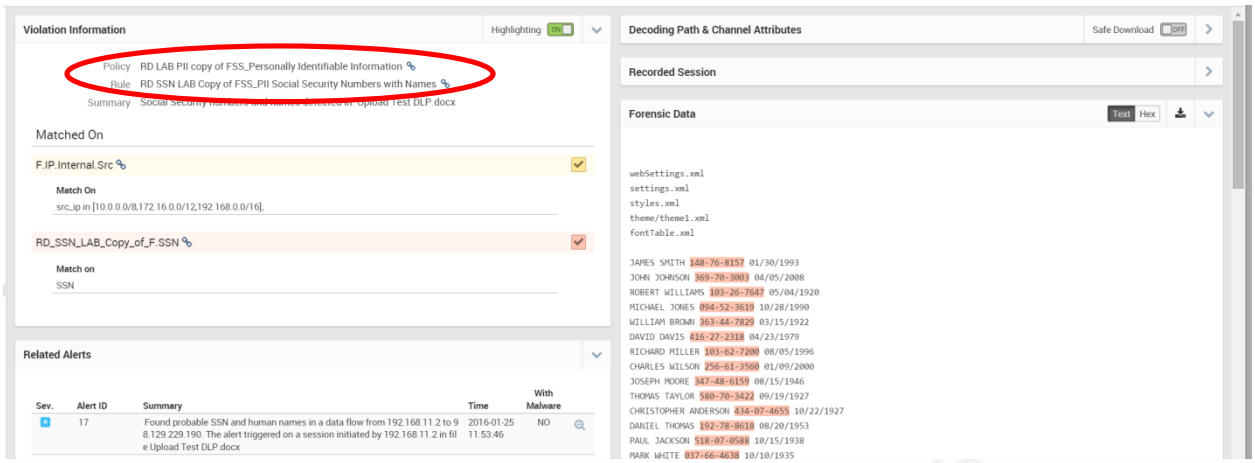

Figure 16. Detailed Alert Response

3.3 Data Exfiltration SSL Decryption Detection

This section adds the use of secure sockets layer (SSL), an encrypted transport protocol, to the previous SPAN/TAP testing scenario. SSL as transport encryption presents a challenge for DPL inspection. Inspection cannot occur without access to the decrypted payload. This testing scenario introduces SSL decryption and forwarding the decrypted SSL traffic to the DLP inspection platform.

The introduction of encrypted communications channel introduces the concept of data exfiltration in overt versus covert methods. Overt exfiltration, attempts to extract data in a manner that is in plain sight. The converse method, covert exfiltration, will attempt to do so in a manner that is not easily detected (Rashid et al., 2014), sometimes in the form of encrypted traffic.

The architecture change is depicted in the following diagram showing the SSL decryption capability in-line with the traffic to be inspected.

© 2016 The SANS Institute Author retains full rights.

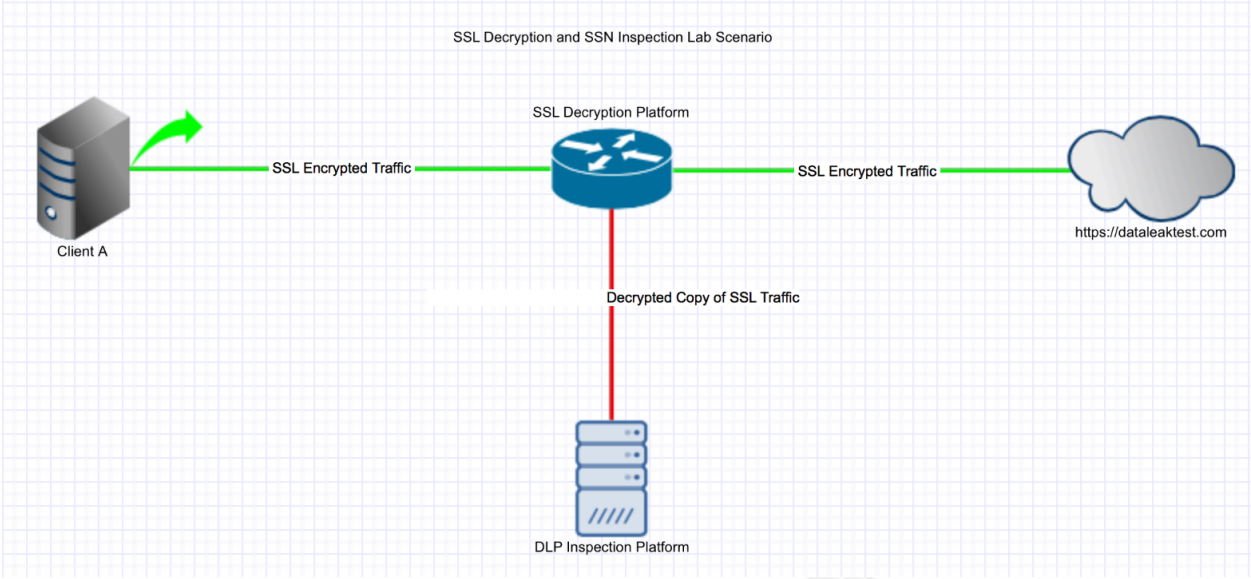

Figure 17. Data Exfiltration SSL Decryption Detection Diagram

The test word document used previously in the SPAN/TAP scenario is attempted to be uploaded to https://dataleaktest.com. This site allows for the encrypted SSL to be used as the protocol.

The alert summary below indicates that SSN's were detected in the file transferred externally. The incident is easily confirmed with the review of alert details. This scenario validates that the DLP out-of-band detection was sent unencrypted data from the SSL decryption and forwarding platform; otherwise, the inspection would not have taken place.

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

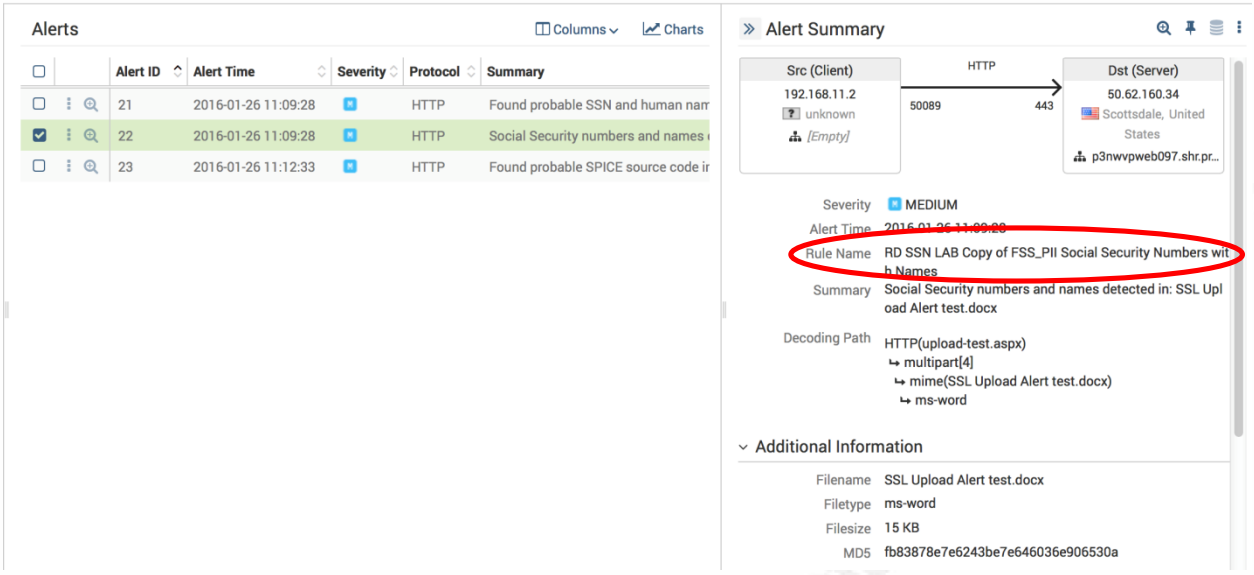

Figure 18. Alert Summary – SSL Visibility (Alert Only)

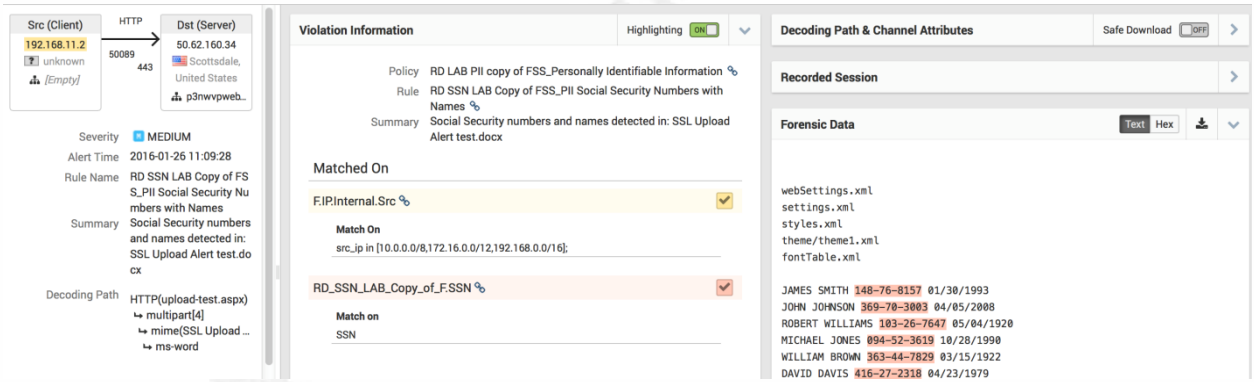

Figure 19. Alert Details

3.4 Data Exfiltration SSL Decryption – Image Detection Bypass

The final scenario leverages all of the capability of the previous SSL decryption and forwarding test. The only change is the word document containing the confirmed valid SSN's has been converted to a portable network graphics (PNG) file. This testing demonstrates the inability for the full DLP Network Based Platform to OCR inspect image files in order to recognize the patterns or content.

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

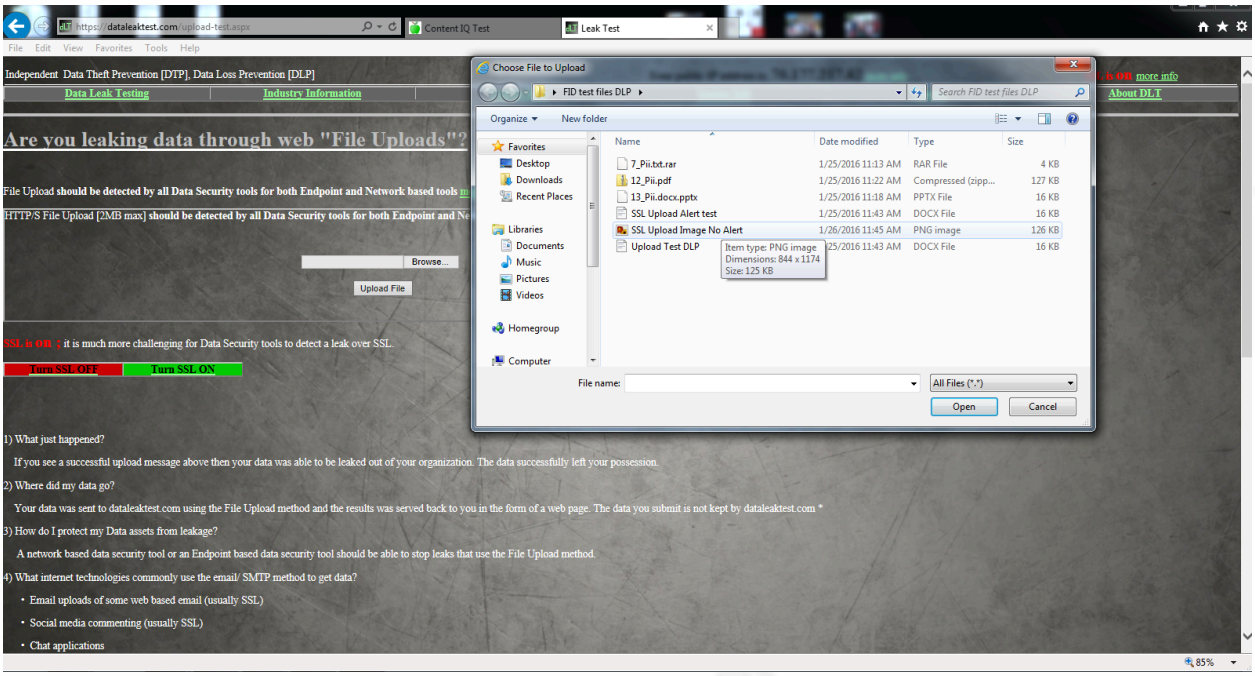

Figure 20. Upload Image

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

JAMES SMITH 148-76-8157 01/30/1993

JOHN JOHNSON 369-70-3003 04/05/2008

ROBERT WILLIAMS 103-26-7647 05/04/1920

MICHAEL JONES 094-52-3619 10/28/1990

WILLIAM BROWN 363-44-7829 03/15/1922

DAVID DAVIS 416-27-2318 04/23/1979

RICHARD MILLER 103-62-7200 08/05/1996

CHARLES WILSON 256-61-3560 01/09/2000

JOSEPH MOORE 347-48-6159 08/15/1946

THOMAS TAYLOR 580-70-3422 09/19/1927

CHRISTOPHER ANDERSON 434-07-4655 10/22/1927

DANIEL THOMAS 192-78-8618 08/20/1953

PAUL JACKSON 518-07-0588 10/15/1938

MARK WHITE 037-66-4638 10/10/1935

DONALD HARRIS 388-06-9327 04/24/1920

GEORGE MARTIN 122-52-5659 10/16/1928

KENNETH THOMPSON 685-05-6808 08/19/1977

STEVEN GARCIA 274-50-1867 07/14/2011

EDWARD MARTINEZ 450-56-5777 11/13/1926

BRIAN ROBINSON 393-54-9040 08/30/1985

Figure 21. PNG File Content

OCR is extremely resource intensive and is not recommended for network-based DLP inspection platforms. OCR is utilized in endpoint and offline DLP inspection platforms. The upload is successful and the evidence of the network traffic exists. A policy violation was not triggered due to the inability to OCR inspect.

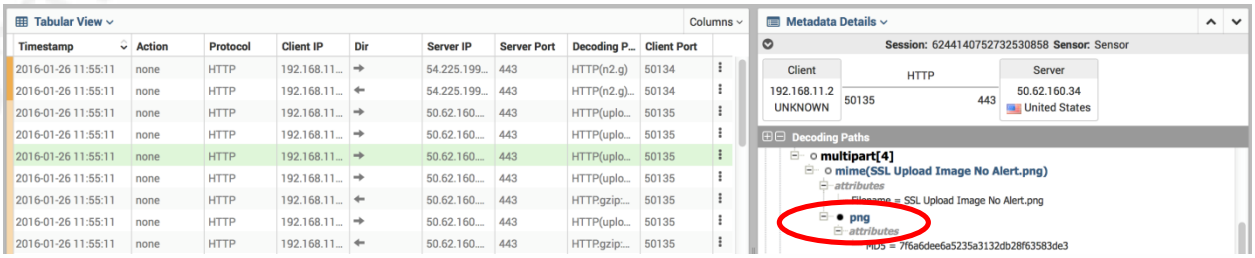

Figure 22. Traffic Detected – No OCR Inspection Capability or Alert

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

In this scenario, the upload attempt successfully transferred sensitive data from within the organization to an external entity.

4. Conclusion

DLP Programs require evaluation, re-evaluation, use-case defining, and testing in order to be successful. Defining and researching methods for DLP visibility, detection and prevention are constant and iterative activities for DLP engineers and analysts. Continual program refinement is driven by the rate at which business systems containing sensitive data are commissioned, the evolution of intricate exfiltration techniques, constant protocol enhancements, and persistent application development. Program refinement challenge the DLP engineer to examine the corporate IT environment, define critical data, and how best to gain visibility of the processing of the critical data.

This research focuses solely on the Data in Motion aspect of DLP, further work is necessary to define and test scenarios. Similar research is required to address Data at Rest and Data on the Endpoint; this research will comprise a complete DLP Protection program for enterprises.

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

References

Basu, E. (2015, October 26). Cybersecurity Lessons Learned From the Ashley Madison Hack

Retrieved from http://www.forbes.com/sites/ericbasu/2015/10/26/cybersecurity-lessons-learned-from-the-ashley-madison-hack/#10fb95bbed99

DLA Piper (2015). Data Protection Laws of the World

Retrieved from http://www.dlapiperdataprotection.com/#handbook/world-map-section/c1_US

Ernst & Young (2011, October). Data loss prevention: Keeping your sensitive data out of the public domain

Retrieved from http://www.ey.com/Publication/vwLUAssets/EY_Data_Loss_Prevention/$FILE/EY_Data_Loss_Prevention.pdf

Finding or Verifying Credit Card Numbers (2013, August 19).

Retrieved from http://www.regular-expressions.info/creditcard.html

Hackett, R. (2015, April 30). Most Internet traffic will be encrypted by year end. Here's why.

Retrieved from http://fortune.com/2015/04/30/netflix-internet-traffic-encrypted/

Kanagasingham, P. (2008, August 15). Data Loss Prevention.

Retrieved from http://www.sans.org/reading-room/whitepapers/dlp/data-loss-prevention-32883

Kumaresan, N. (2014). Key Considerations in Protecting Sensitive Data Leakage Using Data Loss Prevention Tools

Retrieved from http://www.isaca.org/Journal/archives/2014/Volume-1/Pages/Key-Considerations-in-Protecting-Sensitive-Data-Leakage-Using-Data-Loss-Prevention-Tools.aspx

Narayan, K. (2014, October 17). In Plain Sight: How Cyber Criminals Exfiltrate Data Via Video

Retrieved from http://www.darkreading.com/attacks-breaches/in-plain-sight-how-cyber-criminals-exfiltrate-data-via-video-/a/d-id/1316725

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.

Nemschoff, M. (2014, June 28). A Quick Guide to Structured and Unstructured Data

Retrieved from http://www.smartdatacollective.com/michelenemschoff/206391/quick-guide-structured-and-unstructured-data

Ouellet, E. (2013, January 3) Magic Quadrant for Content-Aware Data Loss Prevention

Retrieved from http://www.computerlinks.de/?event=tools.ehgetfile.FileHandler&ting&f_string=/FMS/22876.magic_quadrant_for_content_aware_data_loss_prevent.pdf

Peterson, A. (2014, December 18). The Sony Pictures hack explained

Retrieved from https://www.washingtonpost.com/news/the-switch/wp/2014/12/18/the-sony-pictures-hack-explained/

Rashid, A., Ramdhany, R., Edwards, M., Kibirige, S., Babar, A., Hutchison, D., Chitchyan, R. (2014, April 11) Detecting and Preventing Data Exfiltration

Retrieved from http://www.cpni.gov.uk/documents/publications/2014/2014-04-11-de_lancaster_executive_report.pdf

Sullivan, Dan. (2008). Executive Perspectives on Data Loss Prevention.

Retrieved from www.realtimepublishers.com/chapters/1367/esecdlp-1.pdf

VanAntwerp, R. C. (2011). Exfiltration Techniques: An Examination and Emulation

Retrieved from Thesis: http://udspace.udel.edu/bitstream/handle/19716/10145/Ryan_VanAntwerp_thesis.pdf?sequence=1

Randy Devlin, rdevlin@mastersprogram.sans.edu

© 2016 The SANS Institute Author retains full rights.