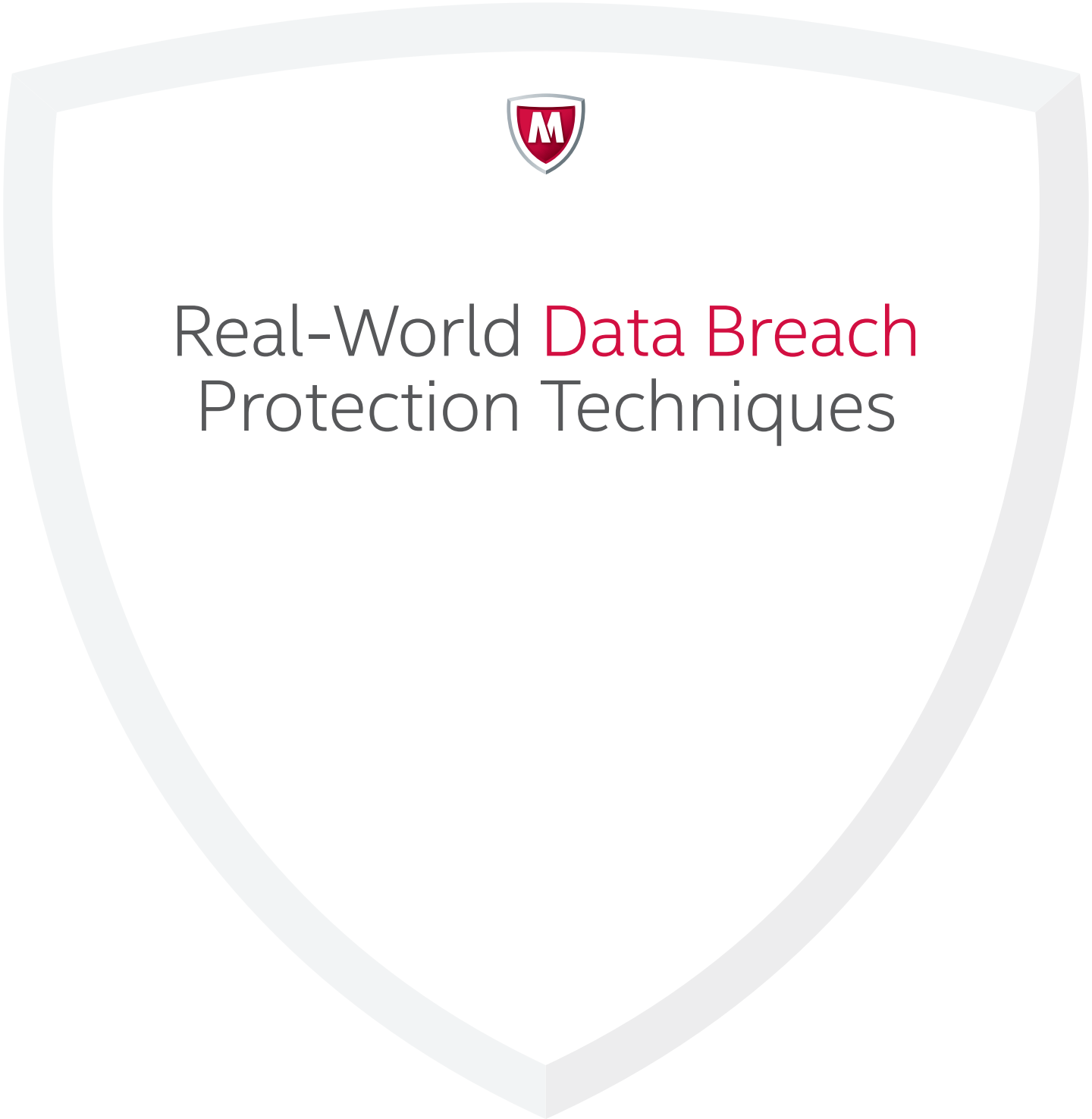
White Paper



Real-World Data Breach Protection Techniques

intel Security

This white paper was written by:
Robert Gresham
Senior Consultant
McAfee® Foundstone®
Professional Services

Table of Contents

Introduction

How do you know when a data breach is occurring? Is it through your intrusion detection system (IDS)/intrusion prevention system (IPS), a new malware analysis tool, a visit from law enforcement, the news media, or employee exit interviews? All of these definitely can notify you, and there are ways to make it harder to hide and greatly reduce data breaches from happening without hampering business missions. By shaping user behaviors and actions into predictable flows, malicious threats will stand out.

Data breaches during the past year have been newsworthy. While some security tools focus on the malware to thwart attackers from getting on your network, this paper is focused on the data, so that you can prevent loss by leveraging advanced defense methodologies. Some say you should treat your network as if you have already lost the war and are compromised. You can agree or disagree, but intrusions are on the rise and so are losses.

We will focus on defining and preventing most of the egress paths, making it difficult for clear text data (and some encrypted avenues) to leave the network without notifying you that your sensitive data just left the building. Wouldn't it be great to have something block threats and notify you that the breach was happening or just happened? Chasing malware is a never-ending battle, zero-day threats show up almost every day, and malware is changing even more frequently. The adversaries have made their way in and are pretending to be your employees, moving through your organization unmonitored. How do we build a moat around our sensitive data? First, we have to know where the data is, what type it is, and how it is used—and then we have to protect it. Knowing where your data is and how it's moving is half the battle. Enter McAfee® Network Data Loss Prevention (McAfee Network DLP).

McAfee NDLP

McAfee Network DLP provides preventive measures to protect data-in-motion, data-at-rest, and data-in-use (when used with McAfee DLP Endpoint). I will focus primarily on data-in-motion, as this is the last line of defense and provides the best notification capabilities without interacting intensively with your systems and assumes that you know where your data is. Otherwise, you should use data-at-rest notifications to identify some of the loose data points and then develop your protection scheme with data-in-motion tools. McAfee Network DLP works in conjunction with McAfee Web Gateway and McAfee Email Gateway. McAfee Network DLP Prevent and McAfee DLP Endpoint are usually used as a "behavioral modification" tools. In simpler terms, it prompts users to fix their issues and correct "accidental" leakages by responding back to originators and requiring them to add justifications or modifications to the original message. This is the true value of McAfee Network DLP, along with meeting compliance and guidance for protection of data. However, it can be used to find rogue leakages and often finds misconfigurations by monitoring the data passing through the network. Let's get to the details of protection.

Data Context

One concept is crucial in helping customers protect their data and it's usually the first lesson of DLP. IT professionals are great at finding content, such as Microsoft Word document or a web post. What IT usually lacks is the context of the data. Why was the data sent this way? Is this a legal use of that data? Does the data viewed provide context about its intent. In the Payment Card Industry (PCI), a credit card number in plain text is bad, but is it really? Or does the credit card number need to have context around it to make it bad? From a presentation perspective, we want to demonstrate that credit card numbers are leaving your network. In most environments, finding credit card numbers and Social Security numbers is easy. However, the context of the credit card numbers is really what we need to focus on. Why? Because a credit card alone is useless without some additional data. Finding the number is easy, finding the context of how that number is used is exponentially

more difficult, especially when you include the use of everyday words with some content like HIPAA's ICD9 or ICD10 definitions. So content is important and the first step in data discovery, but determining the context of the content is crucial and minimizes the endless flow of false positives. Just as intrusion prevention systems (IPS) engineers will modify IPS rules to identify the vulnerability and not the exploit, the exploit will change, but the vulnerability won't. Unfortunately, we don't always have the resources we need to protect the network; we need to automate what we can.

Network Egress Points

Let's look at the Target breach. Research indicates that in the Target breach, data egressed as plain text over FTP. If the data is encrypted or obfuscated, it makes configuration harder and significantly harder to identify naturally. The DLP pundits would say that you could bypass DLP and that it's useless and a waste of time. Like any other controls we have, if there is a will (to bypass/hide/obfuscate), there is a way. To capture encrypted protocols, we'd have to proxy it and we could use SSL intercept for those tools that could support it. Proxies make the malicious use of protocol under encryption, not based on the protocol standards as egress points, harder without notification or capture. Non-standard egress points to non-standard communication points are easy to find if you know your network. Here knowing your network would be crucial to maintaining your protection because allowed non-standard communication ports hide traffic and known, approved end-to-end connections should be always monitored for variance and anomalies.

Umbrella Notification

By placing a McAfee Network DLP Monitor at the boundary of your network at the point just in front of the firewall (recommended) combined with a good aggregation network tap, you have an intrusion detection system (IDS) for your data. Let's call it your exfil detection system (ExDS) to coin a phrase.

A McAfee Network DLP Monitor will passively capture and detect more than 300 content types and match any ASCII character string representation that is defined in a rule. For clarity, McAfee Network DLP cannot ingest encryption or obscured data. Encryption points/paths in your network should always be well defined, known, heavily monitored, and reviewed regularly.

Additional McAfee Network DLP Monitors may be placed for pinpoint monitoring of extremely sensitive data. However, remember this approach won't stop the data leaving—it simply informs you that data has left the building.

Targeted Protection

McAfee Network DLP Prevent supports most mail transfer agents (MTAs) or web proxies as long as they are requests for comments (RFC)-compliant simple mail transfer protocol (SMTP) without transport layer security (TLS) and can interpret message headers or ICAP protocols, respectively. Almost every enterprise has these tools or is planning to have them. We can use them to assist us in channeling the egress routes to our areas of inspection and ultimately provide a denial of exfiltration.

For the sake of clarity, we will not discuss the uses of a mail transfer agent (MTA) and email prevent. The same rule created below could apply to the MTA and emails transitioning from your network with very little effort and can be discussed later.

So let's take our retailer that has web proxy that is compliant with Internet Content Adaptation Protocol (ICAP) standard and has configured it for secure sockets layer (SSL) intercept, excluding banking and finance websites. For obvious reasons, these sites have an Acceptable Use policy for inspection and legal requirements to allow them to inspect traffic on their network. The web proxy is configured and set up to accept connections, decrypt and intercept SSL communications, and forward them on to ICAP port 1344. The web proxy is also set up to filter and proxy FTP, HTTP, HTTPS (with decryption), and act as an Internet relay chat (IRC) proxy.

First, you need to know what traffic is allowed to traverse your firewalls and from what sources. This step helps you successfully identify and block rogue traffic. If you can't discern what good traffic looks like on your network, you will have a hard time identifying rogue traffic within normal production traffic. Knowing your network and egress routes from your sensitive data points is required for successful protection. At times, you may need to adjust the firewall to allow legitimate traffic to go through to meet business requirements.

However, an application-level proxy is our most powerful tool for normalization and one of our best lines of defense against exfiltration. Whether it's a firewall or web application proxy, it can enforce authentication and protocol standards and allow for inspection of payloads beyond simple URL filtering. This is why next-generation firewalls are important. Introducing these new controls will begin to limit exfiltration. This will apply to specific paths and/or processes for getting out of the network or provide specific choke points for normally unprotected traffic. For example, an application proxy can stop and deny non-HTTP-compliant traffic from leaving your environment, a secure shell (SSH) over 443 or netcat tunnel over 443. If this traffic is HTTP-compliant, then it's subject to inspection and capturing. This might seem to be a utopian IT concept in a world where everything works together perfectly. But in the real world, you will have systems bypassing the proxy and firewall rules being aggregated for ease management or for performance. Servers with holes through the firewall allow the server to send data out as it sees fits. McAfee Network DLP Monitor is like your ExDS and will give you statistics for the types of traffic and types of content traffic that is traversing out of your network.

Configuration

So where do you start? First, use an explicit web proxy and send all possible proxy traffic that you can through the web proxy. The proxy is then configured to use ICAP and will forward the requests to the McAfee Network DLP Prevent appliance.

Set up the McAfee Network DLP Web Prevent on the same subnet as the web proxy, and configure the web proxy to send an ICAP REQMOD (request modification) to the Web Prevent. When properly configured and with no rules base, Web Prevent will accept ICAP requests and respond with an HTTP/1.1 200 OK response to the proxy, stating that the data is cleared to POST. Otherwise, a HTTP/1.1 403 Forbidden with ICAP Deny HTML page is presented to the proxy, which is returned to the requestor.

The web proxy should be configured to send at a minimum the HTTP POST requests via ICAP for efficiency. In this scenario, the web proxy is configured to support HTTP POST/PUT, FTP Data, IRC POST requests to the McAfee Network DLP Web Prevent via port 1344 the ICAP standard port.

The firewall is now configured to redirect web, FTP, and IRC traffic directly to the proxy for inspection and all bypasses are configured on the proxies. The firewall is also configured to only allow traffic from the proxy on those protocols. If servers can't negotiate authentication or the proxy, then that traffic is allowed by exception, and extra protection/mitigations should be put in place to make sure that system isn't misused. Proxy interception for other protocols can be handled by the firewall. Soon, firewalls will support ICAP forwarding, and every payload will be checked. In an environment as large as Target's, with its sprawl and management of those firewalls, monitoring egress routes is very important and very time-consuming. This is where McAfee Network DLP Web Prevent may not be the best solution, but McAfee Network DLP Monitor can inspect the outbound traffic and inbound traffic and dissect more than 300 content types to find the ASCII text you are looking for without impacting traffic. You certainly would notice 40 million credit card numbers passing by with McAfee Network DLP and know that this activity does not indicate a false positive.

Now, that we have our McAfee Network DLP Web Prevent properly installed and monitoring traffic, we need to get some rules running on it to spot the data we are trying to protect. McAfee Network DLP works with rules (content with context), concepts (regex keywords), keyword, protocol, and content types to find different types of data in various ways. Concepts are the most powerful tools in the McAfee Network DLP toolkit.

Concepts

Concepts combine content with limited context and allow for very specific matching criteria. Concepts with proximity are a very powerful tool for finding specific types of content around other content. For this example, credit card data requires a few other items to make it useful: customer name, credit card account number, expiration date, CCV2 codes, and other data. Most of this data is contained within the magnetic strip of the card and interpreted by card scanners. Using the Target breach as an example, malware (like BlackPOS) can skim card data from stored memory and then store it in flat files for transport to a consolidation or exfiltration server, much like what occurred during the data breach at Target.

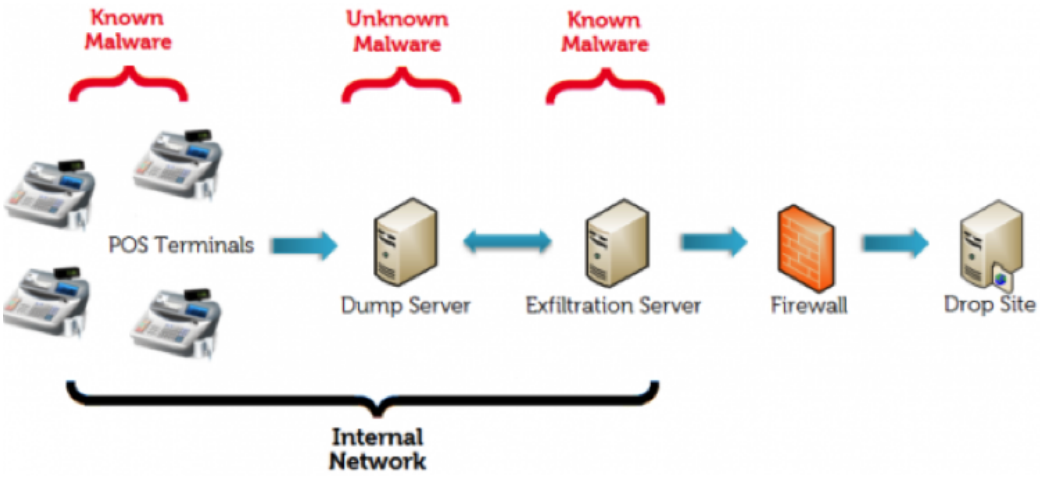


Figure 1. Relationships between compromised and attacker-controlled assets.
(Source: Dell Secureworks and krebsonsecurity.com)

How does McAfee Network DLP work with respect to proximity (or context) and credit card data (or content)? McAfee Network DLP has a built-in concepts that looks for credit card numbers and validates them with the Luhn10 algorithm and regular expressions that are formatted to find specific common credit card account codes in various formats. Let's create a rule that finds these 13- to 16-digit numbers and validates the numbers matching the numbers found by the regular expression against Luhn10 algorithm. Is that enough to be positive that we have a valid credit card number? It casts a wide enough net to generate a false-positive investigation.

Context

Security professionals usually don't have a problem finding content within data streams, as long as they know what they are looking for. Context is more than just the content; it provides viable data usage parameters around the content—the content's intent for use. Finding credit card numbers on your network in a test file wouldn't and shouldn't be a violation because those numbers have no context as to whom they belong or how to use them. However, combining them expiration dates, CCV2s, and customer data gives you a whole new view to the context of the data seen. Next, we will create the context to find skimmed credit card data without the obvious keywords next to them by using context.

First, we need to create a custom concept that will help us find the expiration data off of a magnetic strip. It is in the YYMM format.[1] See our example for McAfee Network DLP below.

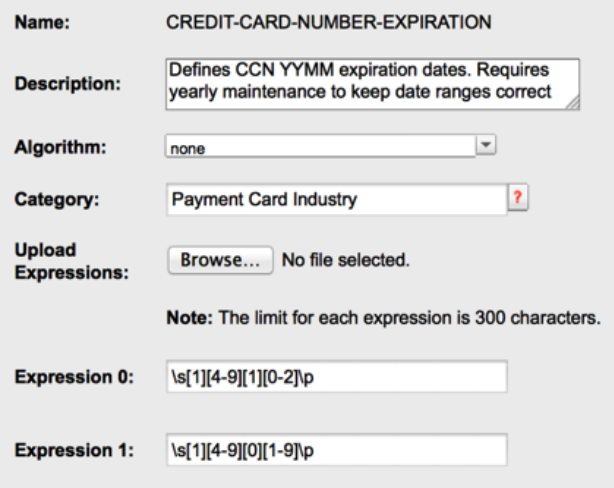


Figure 2. McAfee Network DLP concept for credit card expiration.

The built-in McAfee Network DLP "CREDIT-CARD-NUMBER-GENERAL" under the Payment Card Industry category includes the Luhn10 algorithm check along with 21 regular expressions that look for known credit card formats. We click the "Export Expressions" button and save the file.

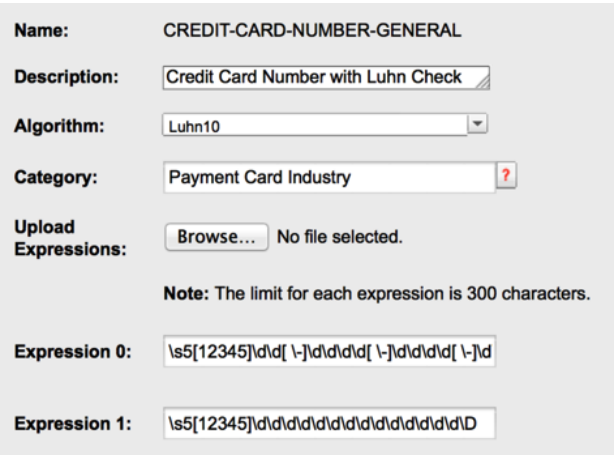


Figure 3. McAfee Network DLP concept for credit card numbers in general. This example does not include all the expressions.

Now, we go back to Payment Card Industry category and create a new concept called "MFE-CREDIT-CARD-EXFIL" and browse to the file we just saved under the "Upload Expressions" button. Enter a description of credit number card check with Luhn10 check in proximity of expiration data, select the algorithm Luhn10, and click "Save." The newly created concept finds numbers matching the regular expressions and then validates them with the Luhn10 algorithm, which checks the number found in three distinct ways to ensure that the number found is not just a random string of 13 to16 digits.[2] Now this doesn't mean it's a real or valid credit card account, but it looks like one and would be worthy of investigation.

We reopen our "MFE-CREDIT-CARD-EXFIL" concept, expand "Count," and enter the values greater than five. Then we go down to the "Proximity" field and select our newly created "CREDIT-CARD-EXPIRATION" concept check within 500 bytes. This concept will validate a suspected credit card number against the Luhn10 algorithm and then look to see if there is number value like YYMM expiration code concept within 500 bytes of the credit card number. This is enough information to find a valid credit card in raw form in a text file. If they parsed the data in any pattern, this configuration would find it. Usually for insider threats, we combine this with a keyword proximity that almost always finds the data accurately. The keyword concept is a reliable way of finding distinct violations that are mostly due to improper procedures or actions on the part of a malicious insider.

Here's what the McAfee Network DLP Credit Card Number (CCN) Data Exfil concept would look.

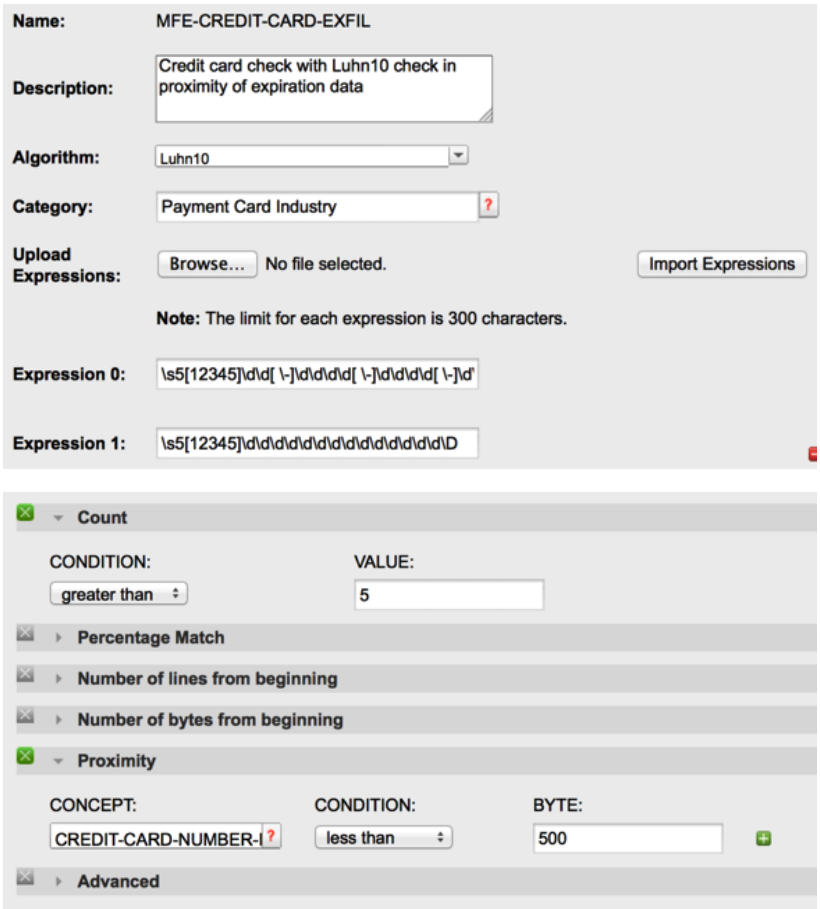


Figure 4. McAfee Network DLP CCN Data Exfil concept.

Finally, we need to put this rule in action. You would apply this concept to a rule within your McAfee Network DLP Web Prevent policy and within your monitor policy. The rule would use the newly created concept and an action rule that would create an incident and notify a group of the traffic, if it is seen. Now we can test it. Remember that monitor action rules and prevent action rules are different: "Monitor is to prevent, and allow is to monitor."

You can use the CSM Test Center at **http://csm-testcenter.org** for a file upload test to ensure your web proxy process is capturing the content we want to see. This can be done with a file containing content below.

The Target breach may have been uncovered as a breach by using the PCI concepts within the McAfee Network DLP, but it would have created a huge amount of alerts. For a critical alert, the count would be above 500 and below 500 for a major alert. Using a security information event management (SIEM) solution would allow you to create correlations of additional value. If you find value in these examples, send us a message and ask for more DLP examples from the field.

About the Author

Rob Gresham, Senior Professional Services Consultant NA, Content and Network Services, Intel Security Group

As a senior professional services consultant on the McAfee Foundstone Professional Services team, Rob is responsible for the McAfee Network Data Loss Prevention service line for North America. He has 15 years of experience in information security—performing information security assessments, security architecture design, risk assessments, data loss prevention, and incident response for multiple healthcare and financial sector customers.

About McAfee Foundstone Professional Services

McAfee Foundstone Professional Services, a division of McAfee, a part of Intel Security, offers expert services and education to help organizations continuously and measurably protect their most important assets from the most critical threats. Through a strategic approach to security, McAfee Foundstone identifies and implements the right balance of technology, people, and process to manage digital risk and leverage security investments more effectively. The company's professional services team consists of recognized security experts and authors with broad security experience with multinational corporations, the public sector, and the US military. **www.foundstone.com**.

About Intel Security

McAfee is now part of Intel Security. With its Security Connected strategy, innovative approach to hardware-enhanced security, and unique Global Threat Intelligence, Intel Security is intensely focused on developing proactive, proven security solutions and services that protect systems, networks, and mobile devices for business and personal use around the world. Intel Security combines the experience and expertise of McAfee with the innovation and proven performance of Intel to make security an essential ingredient in every architecture and on every computing platform. Intel Security's mission is to give everyone the confidence to live and work safely and securely in the digital world. **www.intelsecurity.com**.

1. ISO/IEC 7813:2006, Information technology—Identification cards—Financial transaction cards, **http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43317**
2. ISO/IEC 7812-1:2006, Luhn Algorithm, Identification cards—Identification of issuers —Part 1: Numbering system, **http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=43317**

McAfee. Part of Intel Security.
2821 Mission College Boulevard
Santa Clara, CA 95054
888 847 8766
www.intelsecurity.com