SearchSecurity.co.UK E-Guide

The Basics of Endpoint Security: Expert Reveals Tips for Finding Data on the Endpoints

When an endpoint device goes missing, it isn't the cost of the device that keeps security pros up at night. It's the uncertainty of not knowing what sensitive information was stored on the device. The first step in securing endpoint devices is identifying and eradicating sensitive information stored on them.

In this E-Guide, security expert Mike Chapple reveals tips on finding sensitive information on the endpoint. Get a basic understanding of the algorithms used to detect sensitive information, software tools to assist you in your search for data, and two basic approaches for managing sensitive information.

Sponsored By: PANDA SECURITY

TechTarget
The Technology Media
ROI Experts

SearchSecurity.co.UK E-Guide

The Basics of Endpoint Security: Expert Reveals Tips for Finding Data on the Endpoints

Table of Contents:

How to find sensitive information on the endpoint

Mike Chapple, Contributor

This technical tip is part of SearchSecurity.com's Integration of Networking and Security School lesson, "Back to basics: Endpoint security on a budget." See other materials from this lesson or visit the school homepage for more information.

There's little question that any security manager who's suffered through a lost laptop incident knows what aspect of the ordeal causes an organization real damage. When a laptop goes missing, it's not the loss of a $2,000 asset that causes heartburn; it's the fear, uncertainty and doubt that results from not knowing whether sensitive information was stored on the missing device.

Fortunately, security professionals may take advantage of a number of sensitive information discovery tools to identify and eradicate sensitive information stored on endpoint devices.

Sensitive information discovery algorithms

Before we delve into the tools available to assist with the search, it's important to have a basic understanding of the algorithms used to detect sensitive numbers. Only by understanding how these algorithms work is it possible to judge the effectiveness of individual scanning tools. We'll specifically look at two types of sensitive numbers commonly sought out by sensitive data discovery tools: credit card numbers and Social Security numbers.

Credit card numbers issued by the major providers follow a standard format that makes it easy to detect them using regular expressions. The rules for valid numbers include:

- Visa numbers have either 13 or 16 digits and always start with a 4.
- MasterCard numbers have 16 digits and always start with a 5, followed by a digit between 1-5.
- American Express numbers have fifteen digits beginning with 34 or 37.
- Discover Card numbers have 16 digits beginning with 6011, 622, 644-649 or 65.

These guidelines are a great starting point for ruling out quite a few false positives because they can easily be adapted to a regular expression. For example, the following regular expression can be plugged into a search tool to find potential Visa card numbers, even if there are whitespace characters between the groups of four digits:

```
\b4\d{3}[ -]?\d{4}[ -]?\d{4}[ -]?\d{4}[ -]\b
```

There's also a validation algorithm built into credit card numbers that provides even greater confidence in a match. The Luhn algorithm verifies that a card number passes the "check digit" test, which enables error detection by identifying number patterns that are known to be invalid. The algorithm works by summing all the card number digits and then performing the mod 10 operation on the sum. For those of you forgetting high school math, to perform the mod 10 operation, simply divide the number by 10. The integer remainder is the result. For example, let's verify the following credit card number:

4128 0057 1492 1925

Add the first 15 digits together, which produces a sum of 55. Divide that by 10 and you get "5 remainder 5". In other words, the remainder (5) matches the last digit of the card number (also 5), so you know it is potentially valid. This algorithm does not, of course, confirm the number corresponds to an active account, but it does provide additional confidence in your match.

Social Security numbers, on the other hand, are not quite as easy to match because there is no Luhn algorithm equivalent to verify their validity. You can look for patterns of nine digit numbers surrounded by white space and take advantage of a few clues to help with the search:

- SSNs are often (but not always) written with hyphens between the digits, in the form xxx-xx-xxxx. If you're willing to accept the risk of missing unformatted numbers, you can restrict your search to numbers hyphenated in this pattern to dramatically reduce false positives.
- SSNs will never have all 0's in a digit group (i.e. 000-xx-xxxx, xxx-00-xxxx or xxx-xx-0000).
- SSNs will never begin with 666, 732-749 or any number higher than 772.
- Given the first three digits of an SSN, you can determine the highest possible values for the next two digits by consulting the Social Security Administration's High Group Number list.

Software tools to assist in the search

Unless you're looking for an adventure, it's not necessary to write your own code to implement these searches. There are a variety of open source and commercial products available to assist you in detecting these sensitive numbers on enterprise systems. Some examples include:

- Cornell University's Spider
- University of Texas at San Antonio's Sensitive Number Finder
- Identity Finder

These tools use the algorithms described above and allow you to tinker with the settings, such as whether to restrict a search to formatted numbers, numbers in particular file types and other parameters.

Managing sensitive information and data

After deciding upon a search strategy for finding potentially sensitive information, the next step is to decide on a strategy for managing the mountains of results data.

There are two basic approaches to this problem: centralized review or decentralized authority. In the centralized approach, the tools report all results to a central administrator who is responsible for validating and eradicating suspicious data. This is an extremely time-consuming process and taxes valuable IT resources. However, it ensures consistency of rule interpretation and the thorough review of findings.

In the decentralized approach, end users are given responsibility (and accountability) for reviewing results. This distributes the workload among the entire workforce and provides the added benefit of having staff with contextual knowledge perform the review. For example, a staffer who knows that an Excel spreadsheet contains information about parts orders may be able to immediately disregard reports of SSNs in that document, while a centralized reviewer might not know the difference between that and any other file.

The downside of this approach is obvious: It's a lot harder to get all the individuals in an organization to search their systems than it is to have a centralized staff perform the searches as a core responsibility. If you choose the decentralized approach, you'll probably want to develop a reporting mechanism to allow individual employees to provide progress reports, allowing you to track down those that ignore your requests. You'll also want to provide detailed training, perhaps through the use of screencast videos or detailed documentation, walking users through the scanning process. Finally, you'll need to make technical support available to users who have difficulty performing or interpreting the scans.

Scanning systems for sensitive data is a complex problem but, fortunately, there are a variety of tools and techniques available to assist in the process. Minimization, the searching and eradication of sensitive information on endpoints, is a powerful strategy in the arsenal of security administrators seeking to reduce enterprise risk.

About the author:
Mike Chapple, CISA, CISSP, is an IT security professional with the University of Notre Dame. He previously served as an information security researcher with the National Security Agency and the U.S. Air Force. Mike is a frequent contributor to SearchSecurity.com, a technical editor for Information Security *magazine and the author of several information security titles, including the* CISSP Prep Guide *and* Information Security Illuminated. *He also answers your questions on network security.*

timefor yourbusiness

After the MEETING, a COFFEE break.

"Meetings, working lunches, dealing with suppliers… After all this, I need a break. At least I have peace of mind knowing I can leave the security of my systems in the hands of Panda Security."

Lee Adams, 38. IT manager. *A Secure Businessman.*

One worry less. Time for your business.

PANDA GLOBAL BUSINESS PROTECTION

SECURITY SOLUTIONS FOR YOUR COMPANY

Whatever the size of your organization, from the smallest business to the largest corporation, **Panda Security provides the global solution that best adapts to your protection needs.** Easy to install, easy to maintain and with a minimum investment.

What's more, our products provide **maximum protection through Collective Intelligence**, disinfecting thousands of new threats every day, including those not detected by other security solutions.

MAXIMUM DETECTION · MINIMUM CONSUMPTION · COLLECTIVE INTELLIGENCE

COMPLETE PROTECTION FOR:

- Email
- Web traffic
- PCs and servers

Panda Security lets you get on with your business.
Find out more at **www.pandasecurity.com/timeforyourbusiness**

PANDA SECURITY | 20th Anniversary 1990-2010

Resources from Panda Security

Protect Your Company, Freeing Up Time for Your Business

Switching from Anti-virus to Software-as-a-Service (SaaS)

Panda Managed Office Protection FREE DEMO

About Panda Security
Founded in 1990, Panda Security is the world's leading provider of cloud-based security solutions with products available in more than 23 languages and millions of users located in 195 countries around the world. Panda Security was the first IT security company to harness the power of cloud computing with its Collective Intelligence technology. This innovative security model can automatically analyze and classify thousands of new malware samples per day, guaranteeing corporate customers and home users the most effective protection against Internet threats with minimum impact on PC performance. Panda Security has 56 offices throughout the globe with US headquarters in California and European headquarters in Spain.

Panda Security collaborates with Special Olympics, WWF and Invest for Children as part of its Corporate Social Responsibility policy.