Guide

coresecurity
by HelpSystems

Advanced Threat Detection for the 21st Century

Integrating Big Data Intelligence to Enable Breach Prevention

Integrating Big Data Intelligence to Automate Breach Defense

Today's IT organizations are at a significant disadvantage when it comes to protecting their businesses against advanced malware. These attacks are taking place on an uneven battlefield, with the balance of power skewed in the threat actor's favor. First, the threat actors have the first move. Second, they know more about you than you know about them. For example, they have the resources to find out if you are running a sandbox and they have the tools to build malware that evades that sandbox.

Technology companies have been building solutions to try and even the playing field for the past ten years. While layered security builds a robust safeguard against attack, threat actors are constantly trying new tactics, and advanced malware is getting through. Despite their best efforts, IT organizations are often still at risk of breach.

How do we live in a compromised world? How do we catch infections before they result in a data breach? Answering these questions is Core Security's mission. Using big data science, we work to outmaneuver attackers, preventing and reducing the risk of infections and data breaches that can devastate organizations.

Understanding the Threat Lifecycle

Before we describe how Core Security uses big data to detect threats that get past your defenses, it is important to understand how advanced malware works. The threat lifecycle often begins with a file dropped on a machine. The drop can happen anywhere—at your office, a coffee shop, or the user's home. And it can occur any number of ways—a drive-by attack, an automatic download from a compromised website, etc. Once on the machine, the dropper communicates externally to receive an update and become available for missions as determined by the threat actor. Every time the dropper gets a new mission it is updated and the binaries are changed. The new binaries are an encrypted payload, so if you don't see the file at the beginning of the lifecycle, you won't see the malware at all.

Advanced malware is often defined as the ability to evade defenses by changing rapidly. Changes occur in the malware binaries themselves as well as the destinations of the Command and Control (C&C) servers, but the one sole element that is fundamental is the consistent need for an infected device to communicate over a network to the threat actor. That requirement to communicate is also the Achilles heel of any attack.

A majority of these communications between the malware and C&C servers appear as regular HTTP web traffic, utilizing the Domain Name System (DNS) protocol. Why DNS? Threat actors want agility, availability, reliability, and anonymity in their infrastructure, just like anyone else. As the critical building blocks of the Internet, domain names, and DNS are the only answer. If threat actors limited their infrastructure to a single IP address, it would be easy to render it ineffective by adding the infrastructure to a blacklist. It's harder to swap around IP addresses than it is to get new domain names, which threat actors do in bulk on a daily basis. At the end of the day, DNS provides ability, availability, reliability, and a higher level of anonymity for the threat actor. As a result, if you can limit DNS abuse, you can limit the overall abuse on the Internet.

Building a Breach Prevention Platform: Finding the Needle in the Alerts Haystack

Core Network Insight is a network security appliance that applies engines to the network traffic coming through to look for evidence of infection. If there's enough evidence of an infection, the appliance sends confirmation and definitive evidence of the breach, allowing you to begin remediation efforts immediately.

The key here is not to look for a single indicator of compromise, but instead to look at all network communications and behaviors across the entire threat lifecycle to see if there are any indicators of compromise. For example, if we do not get to see the actual file coming across the network because the asset became infected while it was off the corporate network, then we see communications that occur later in the lifecycle that suggest this is an infected asset. Perhaps the asset communicates back to a site on the Internet that we know is bad or uses peer-to-peer (P2P) or other types of communication on the network that are suspicious.

The beauty of this approach is that it can allow Core Security to catch threats months ahead of anyone having a signature for the actual malware. Instead of modeling the actual malware, Core Security models the communication procedures of the threat operators themselves. Once you start modeling the threat operators, it doesn't matter what malware threat actors build—this now allows us to see their behavior. This allows us to know how they operate so you can identify new threats without ever having seen the malware itself.

Leveraging Big Data: the Core Security Secret Formula

A simple formula describes Core Network Insight, and it begins with big data. Core Security takes in 22.5 billion records of Passive DNS data everyday from various sources. We see a tremendous amount of data—about 43% of North America's wired Passive DNS data and about 1/3 of the mobile data traffic. Passive DNS is important because it's hierarchical, and it tells us IP address to domain name pairings that allow us to see where devices are going on the Internet.

Core Security takes this big data, and we apply machine learning to it. Our data scientists look at the data to find features indicative of an infection. By using machine learning, our data scientists can build classifiers that allow them to automatically identify malicious network traffic versus benign traffic. These classifiers allow the data scientists to build profilers—detection engines built into Network Insight and updated in real-time based on a continued analysis of big data with our classifiers on the back end and inside Core Security. We have 12 of these detection engines in Network Insight.

Below we profile three of these, which harness big data to spot threats before they are clearly visible to the broader security community.

Profiler: HTTP Request Profiling

Malware uses HTTP to "blend in" and evade detection by sending small traces of information over the core ports and protocols that enterprises allow in and out of their network. Leveraging Core Security's big data harvesting and machine learning systems, the HTTP request profiler within Core Network Insight can statistically identify similar structures within HTTP requests to discover hidden infected devices. In customer trials, the HTTP request profiler detected five times the number of active infections that traditional technologies found.

Profiler: P2P Profiling

Another example of how Core Security harnesses big data to analyze emerging forms of malware communication is our peer- to-peer (P2P) profiler. As malware continues to evolve, much of the most up-to-date malware—including ZeroAccess, TDL v4, and Zeus v3—are now leveraging P2P capabilities to evade detection from traditional signature, sandboxing, and blacklisting techniques. Leveraging our big data set, Core Security has built classifiers that allow us, in real time, to look at P2P traffic and identify specific malware families communicating back to C&C to get instructions and updates.

Core Network Insight performs flow analysis on egress traffic and uses machine-learning algorithms to classify the traffic associated with P2P swarms as benign or malicious command-and- control traffic and pinpoint which endpoints are infected.

Profiler: Domain Fluxing Profiler

Our key profiler leveraging DNS is the domain fluxing profiler, which enables the proactive detection of DGA-based botnets. Domain-generating algorithms (DGAs) are techniques used by advanced malware to evade common detection and prevention mechanisms.

Both the malware and the threat actor look up a seed everyday. They take the seed and feed it into an algorithm that generates a random set of domains—as many as 1,000 every day. The threat actor selects one domain and registers it. The malware looks up all 1,000 domains until it finds the one the threat actor registered, then communicates with it and gets the information it needs. There are more DGA based botnets than in the past, and DGA is often use as a fallback communication method for callback to a C&C infrastructure.

Network Insight in Action: The DGA profiler and Pushdo

Traditionally, security researchers deal with DGA-based botnets by reverse engineering the malware to come up with the DGA. But this approach doesn't scale because the malware can be updated every step of the way, forcing security researchers to go through the process of reverse engineering the malware all over again.

Core Security's DGA Profiler takes a different approach. It can take data in the network and automatically identify DGAs in play. Every time you look up a domain and the DNS server can't find it, it generates an NXDomain (non-existent domain) record consisting of the domain name and the IP address looking it up. Based on our machine learning of NXDomains, we've identified features such as the length, level of randomness, character frequency, and domain structure, which tell us this is a DGA.

Using DGA classifiers we were able to detect a new variant of the malware PushDo a full six weeks before most of the antivirus communities had a signature for it. We began detecting malicious traffic before anyone saw the malware. Other examples of discoveries leveraging our DGA Profiler include a new iteration of the TDSS/TDL4 malware and the Mac Flashback virus.

Detect with Certainty Using Big Data

Network Insight is built with scientific research and big data visibility, automatically and accurately identifying hidden infections in real time on live traffic. Network Insight has a decade worth of evidence collected from millions of devices worldwide, providing both new and historical insights that can help verify infections before they can do damage. Using big data and sophisticated automation, we can finally begin to keep up with today's evolving threats.

coresecurity
by HelpSystems

www.coresecurity.com

About HelpSystems

HelpSystems is a people-first software company focused on helping exceptional organizations Build a Better IT™. Our holistic suite of security and automation solutions create a simpler, smarter, and more powerful IT. With customers in over 100 countries and across all industries, organizations everywhere trust HelpSystems to provide peace of mind. Learn more at www.helpsystems.com.