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, and the balance of power is 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 that try to prevent these attacks for the past ten years. And IT organizations have faithfully deployed them, layering security control over security control in an attempt to deploy defense-in-depth. But our defenses further skew the battlefield, adding costs and complexity while at the same time still not preventing the attacks. IT organizations and the threat defense market are realizing that prevention is never 100%; in fact, it's much less than 100%. Advanced malware is getting through these traditional defenses, leaving IT organizations with a risk of breach.

In an effort to close this security gap, organizations are attempting to mitigate the risk of infection with human resources – having people scour event logs to identify malicious activity on the network. But this, too, puts IT organizations at a disadvantage because it is very difficult to find trained humans that can do this work and once you do, it simply takes too long to respond to threats once they're in the network. Meanwhile, IT organizations are being pressured by lines of business and management to be more innovative and adopt technologies like cloud computing and BYOD that further increase the business' risk posture.

Clearly, if we are to even out the battlefield, organizations need a new approach to security. Infections will happen; malware will come through. The question becomes, How do we live in a comprised world? Or How do we catch infections before they result in a data breach? This is Core Security's mission: to lower the risk of data breach. We work to prevent and reduce the risk of the breach as a result of infections that get through your preventative defense. As you'll see, the answer lies not in malware analysis but in big data science.

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 re-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 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 the 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. That's what Core Security seeks to do.

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 convicts the asset and informs you of the infection.

The key here is not to look for any single indicator of compromise, but instead to look at all network communications and behaviors across the entire threat lifecycle to see if any are an indicator of compromise. So, 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 indicate 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 indicative of an infection.

The beauty of this approach is that it often allows 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, and if you see that behavior, 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 thedata 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 nine 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

HTTP continues to be the predominant channel, used by 80% of all malware to communicate. In addition, over 75% of this malware evades detection by traditional protection methods. It is using 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 recent 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 traffic 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. We see more DGA-based botnets than in the past, and DGA is often used as a fallback communication method for call-back to a C&C infrastructure.

Here's how DGA works: 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.

Everyday the whole process begins again.

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 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. Once we got the malware, we sinkholed some of the domain names and learned a lot about it before it was determined to be a new variant of PushDo. Other examples of discoveries leveraging our DGA Profiler include a new iteration of the TDSS/TDL4 malware, and the Mac Flashback virus.

ABOUT CORE SECURITY

Core Security provides market-leading, threat-aware, identity, access and vulnerability management solutions that provide actionable intelligence and context needed to manage security risks across the enterprise. Solutions include multi-factor authentication, provisioning, Identity Governance and Administration (IGA), Identity and Access Intelligence (IAI), and Vulnerability Management (VM). The combination of these solutions provides context and shared intelligence through analytics, giving customers a more comprehensive view of their security posture so they can make better security remediation decisions and maintain compliance.

Core Security is headquartered in the USA with offices and operations in South America, Europe, Middle East and Asia. To learn more, contact Core Security at (678) 304-4500 or info@coresecurity.com.

blog.coresecurity.com | p: (678) 304-4500 | info@coresecurity.com | www.coresecurity.com

CORE SECURITY